***** MANUSCRIPT IN PREP FOR GLOBAL BIOGEOCHEMICAL CYCLES *****

How well do global ocean biogeochemistry models simulate dissolved iron distributions?

Alessandro Tagliabue[1], Olivier Aumont[2], Ros DeAth[3], John P. Dunne[4], Stephanie Dutkiewicz[5], Eric Galbraith[6], Kazuhiro Misumi[7], J. Keith Moore[8], Andy Ridgwell[3,9], Elliot Sherman[8], Charles Stock[4], Marcello Vichi[10,11], Christoph Völker[12] and Andrew Yool[13]

1. School of Environmental Sciences, University of Liverpool, Liverpool, UK
2. IRD-LOCEAN, Institut Pierre Simon LaPlace, 4 Place Jussieu, 75252 Paris Cedex 05, France.
3. School of Geographical Sciences, University of Bristol, Bristol, UK.
4. NOAA/Geophysical Fluid Dynamics Laboratory, 201 Forrestal Road, Princeton NJ 08540, USA
5. Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, MA 20139. USA.
6. ICREA-ICTA, Universitat Autonoma de Barcelona, Bellaterra, Barcelona, Spain
7. Environmental Science Research Laboratory, Central Research Institute of Electric Power Industry, Abiko, Chiba, Japan.
8. Department of Earth System Science, University of California, Irvine, CA. USA.
9. Department of Earth Sciences, University of California, Riverside, CA 92521, USA
10. Department of Oceanography, University of Cape Town, Cape Town, South Africa
11. Nansen-Tutu Centre for Marine Environmental Research, Cape Town, South Africa
12. Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
13. National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton, SO14 3ZH, UK

Main Point 1: First intercomparison of 13 global iron models highlights key challenges in reproducing new iron data
Main Point 2: Wide uncertainty in iron input fluxes, which results in poorly constrained residence times
Main Point 3: Reducing uncertainty in scavenging and biological cycling is a priority

Abstract

Numerical models of ocean biogeochemistry are relied upon to make projections about the impact of climate change on marine resources and test hypotheses regarding the drivers of past changes in climate and ecosystems. In large areas of the ocean, iron availability regulates the functioning of marine ecosystems and hence the ocean carbon cycle. Accordingly, our ability to quantify the drivers and impacts of fluctuations in ocean ecosystems and carbon cycling in space and time relies on first achieving an appropriate representation of the modern marine iron cycle in models. When the iron distributions from thirteen global ocean biogeochemistry models are compared against the latest oceanic sections from the GEOTRACES programme we find that all models struggle to reproduce many aspects of the observed spatial patterns. Models that reflect the emerging evidence for multiple iron sources or subtleties of its internal cycling perform much better in capturing observed features than their simpler contemporaries, particularly in the ocean interior. We show that the substantial uncertainty in the input fluxes of iron results in a very wide range of residence times across models, which has implications for the response of ecosystems and global carbon cycling to perturbations. Given this large uncertainty, iron-fertilisation experiments based on any single

current generation model should be interpreted with caution. Improvements to how such models represent iron scavenging and also biological cycling are needed to raise confidence in their projections of global biogeochemical change in the ocean.

1. Introduction

With the important role played by dissolved iron (DFe) in regulating ocean biogeochemical cycles well established [*Boyd and Ellwood*, 2010], most three dimensional global biogeochemistry models now include a prognostic DFe tracer as standard. These models explicitly represent the DFe limitation of primary production that is prevalent across large areas of the ocean [*C M Moore et al.*, 2013]. This has allowed quantitative projections regarding the impacts of environmental change in Fe-limited regions [*Bopp et al.*, 2013], how DFe may regulate glacial-interglacial changes to the global carbon cycle [*Tagliabue et al.*, 2009] and the wider role played by different nutrients as drivers of planktonic diversity [*Ward et al.*, 2013]. However, the robustness of these results is reliant on how a given model represents the ocean DFe cycle. For example, a model that accounted for hydrothermal sources of Fe was shown to be less sensitive to changes in aeolian iron supply than the same model without a hydrothermal input [*Tagliabue et al.*, 2010]. Equally, there is a six-fold difference in the estimated impact of dust variations on glacial and interglacial changes in atmospheric CO_2 (5-28 ppm) [*Kohfeld and Ridgwell*, 2009] that is largely driven by details of the modeled DFe cycle.

In brief, the ocean iron cycle is regulated by a complex array of different processes [*Boyd and Ellwood*, 2010]. DFe is thought to be supplied to the ocean from atmospheric deposition [*Jickells et al.*, 2005], continental margins [*Elrod et al.*, 2004] and hydrothermal vents [*Tagliabue et al.*, 2010], with potential emerging roles for input from rivers [*Rijkenberg et al.*, 2014], icebergs [*Raiswell et al.*, 2008] and glaciers [*Gerringa et al.*, 2012]. DFe is relatively insoluble in oxygenated seawater and DFe levels are maintained to a large part due to complexation with organic ligands that bind Fe [*Gledhill and Buck*, 2012]. Unbound, or free Fe can then precipitate as solid forms or be scavenged by particles [*Bruland et al.*, 2014]. DFe is operationally defined by the filter size (usually 0.2μm) and over half of the DFe pool can be colloidal [*Boye et al.*, 2010; *Fitzsimmons and Boyle*, 2014; *Wu et al.*, 2001]. This implies that the aggregation and coagulation of colloidal Fe, termed 'colloidal pumping' [*Honeyman and Santschi*, 1989], may also be an important loss of DFe. As a divalent metal, Fe also undergoes rapid redox transformations between Fe(II) and Fe(III) species mediated by oxidation, reduction and photochemical processes [*Wells et al.*, 1995]. The biological cycling of Fe is also complex with varying cellular requirements for Fe [*Raven*, 1988; *Raven et al.*, 1999] and the role of luxury uptake [*Marchetti et al.*, 2009] driving a wide range in phytoplankton Fe quotas [*Sunda and Huntsman*, 1997; *Twining and Baines*, 2013]. Equally, the recycling of DFe by bacteria, viruses and zooplankton is emerging as a key component in governing the Fe supply to phytoplankton [*Barbeau et al.*, 1996; *Boyd et al.*, 2012; *Hutchins and Bruland*, 1994; *Strzepek et al.*, 2005]. Lastly, process studies and basin scale data syntheses have highlighted important specificities to the remineralisation lengths scale and vertical profile of DFe, relative to other nutrients [*Frew et al.*, 2006; *Tagliabue et al.*, 2014c; *Twining et al.*, 2014].

The earliest global iron models were informed by the first efforts to synthesise the emerging datasets on DFe in the late 1990s [*Johnson et al.*, 1997]. These models only considered a dust source, applied constant phytoplankton Fe demands and inferred that the seemingly constant deep ocean DFe concentrations indicated a threshold stabilisation of DFe by organic ligands [*Archer and Johnson*, 2000; *Lefèvre and Watson*, 1999]. As available DFe datasets expanded, it became clear that deep ocean concentrations were more regionally and temporally varied

than accounted for by these models and that explicitly computing un-complexed DFe led to a better model-data agreement [*Parekh et al.*, 2004]. At the same time, assumptions regarding fixed iron solubility in dust and constant C:Fe ratios in exported organic matter were being questioned and alternatives tested [*Ridgwell*, 2001; *Watson et al.*, 2000]. Towards the end of the Joint Global Ocean Flux Study (JGOFS) era more complicated treatments of the demand for DFe from different phytoplankton groups also emerged and when coupled to realistic models of ocean circulation, provided the first estimates of the areal extent of DFe limitation [*Aumont et al.*, 2003; *Moore et al.*, 2002]. In more recent years, and particularly with the advent of the GEOTRACES programme (www.geotraces.org), observations of DFe have expanded rapidly [*Mawji et al.*, 2015; *Tagliabue et al.*, 2012]. This has driven the representation of DFe sources associated with margin sediments [*Moore and Braucher*, 2008] and hydrothermal vents [*Tagliabue et al.*, 2010] in models. At the same time efforts to account for redox speciation [*Tagliabue and Völker*, 2011] and variability in Fe binding ligands [*Misumi et al.*, 2013; *Völker and Tagliabue*, 2015] in global models have also been undertaken.

Until now there has been no comprehensive effort to evaluate how different global models represent DFe, apart from the one off model-data comparisons typical of individual publications [*Moore and Braucher*, 2008; *Tagliabue et al.*, 2008]. Our maturing vision of the oceanic distribution of DFe and our deeper understanding of how it interacts with broader biogeochemical cycles now allows a more widespread intercomparison of global iron models. In conducting the first 'iron model intercomparison project' (FeMIP) we aim to intercompare as broad a suite as possible of global ocean biogeochemistry models with a focus on the reproduction of features present in the full depth ocean sections emerging from the GEOTRACES programme. In doing so we highlight the challenges present for global ocean biogeochemistry models in simulating the distribution of DFe, which emerges as unique to that of other nutrients.

2. Methodology

2.1 Intercomparison process

The goal of this study was to include as many global iron models as possible in order to ensure a 'state of the art' view on their representation of Fe cycling. In that regard, our thirteen models (Table 1) range from those used in the recent IPCC report for coupled climate-carbon studies, to those focused on global patterns of Fe cycling and effects on ocean biogeochemical cycles and phytoplankton diversity, to those concerned with geological timescales. This inclusive design thus did not impose a rigid set of guidelines regarding the model forcings, as done for the ocean carbon-cycle model intercomparison (OCMIP) and climate model intercomparison (CMIP) projects. While imposing identical ocean circulation or external forcing scenarios would have permitted a more direct cross comparison of the different iron models, the extra constraints would have drastically reduced the number of Fe models able to participate and hinder our aim to account for the full diversity of Fe models. Groups submitted their best representation of the dissolved iron distribution in netCDF format at monthly frequency for a canonical year on their standard model grid, alongside additional requested information (temperature, salinity, nitrate, phosphate and silicic acid concentrations, where available). We compiled model data from thirteen model configurations: BEC [*J K Moore et al.*, 2013], BFM [*Vichi et al.*, 2007], BLINGv0 [*Galbraith et al.*, 2010], COBALT [*Stock et al.*, 2014], GENIE (Fe scheme as summarised by [*Matsumoto et al.*, 2013]), MEDUSA1 [*Yool et al.*, 2011], MEDUSA2 [*Yool et al.*, 2013], MITecco [*Dutkiewicz et al.*, 2015], MITigsm [*Dutkiewicz et al.*, 2014], PISCES1 [*Aumont et al.*, 2015], PISCES2 [*Resing et al.*, 2015; *Völker and Tagliabue*, 2015], REcoM [*Hauck et al.*, 2013] and TOPAZ [*Dunne et al.*,

2013], all implemented at the global scale. All models were then regridded onto a 1° x 1° horizontal grid with 33 vertical levels (bounded by 0, 10, 20, 30, 40, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2500, 3000, 3500, 4000, 4500, 5000 and 5500m) as a common FeMIP grid.

2.2 Observational datasets

Observations of dissolved iron are taken from two sources. Firstly, we use an updated version of a global DFe database [*Tagliabue et al.*, 2012] with approximately 20,000 individual observations. This database was gridded at monthly resolution on the FeMIP grid to compare models and observations grid cell by grid cell and month by month, with no volume weighting. Secondly, we extracted DFe data from recent GEOTRACES sections from the 2014 intermediate data product [*Mawji et al.*, 2015]. For comparison purposes (Sec 3.2) the modeled DFe from the longitude, latitude and month of each sampling station was then extracted and the observed data was regridded on the same 33 vertical levels as the models (averaging where more than one observation was present in a particular depth bin). We use datasets collected on the GA-02 West Atlantic cruise [*Rijkenberg et al.*, 2014], the GA-03 North Atlantic zonal transect [*Hatta et al.*, 2014], the CoFeMUG south Atlantic zonal cruise [*Saito et al.*, 2013], the GIPY-6 Atlantic sector of the Southern Ocean cruise [*Chever et al.*, 2010; *Klunder et al.*, 2011] and the recently completed GP-16 Equatorial Pacific zonal section [*Resing et al.*, 2015] that is not yet in the GEOTRACES data product. We note that all IDP2014 GEOTRACES data [*Mawji et al.*, 2015] is also included in the global dataset.

2.3 Brief introduction of the different iron models

The goal here is not to exhaustively describe the FeMIP models for which we refer to the original publications. Rather we seek to summarise how the models treat key components of the Fe cycle and to highlight important differences (Table 1). In our summary we focused on how each model treated the sources of Fe, the chemistry of Fe (including the representation of Fe binding ligands, how free Fe is computed and whether scavenging is a first order rate or a second order function of particle concentrations), biological cycling of Fe (if Fe/C ratios were variable and if zooplankton excretion of Fe depends on the Fe content of prey) and particle Fe dynamics (how many particle pools were simulated and whether the Fe regeneration efficiency was unique or coupled to organic matter).

All models considered a dust source of Fe and only BFM, GENIE and MEDUSA1 did not consider sedimentary Fe supply, only BEC, BFM, PISCES1 and PISCES1 include river input of Fe, while BEC and PISCES1 and PISCES2 are the only models that represent hydrothermal Fe input. All models except BEC compute the free Fe concentration that can be scavenged based on *Parekh et al.* [2004] and all except BFM, COBALT, MEDUSA1 and MEDUSA2 have a second order scavenging rate, i.e. a dependency on particle concentrations. Only PISCES1 and PISCES2 include a representation of colloidal losses of dFe, based on aggregation of dissolved organic material [*Aumont et al.*, 2015]. It is notable that despite a maturing understanding of the variations in the concentrations of Fe binding ligands [*Gledhill and Buck*, 2012], most FeMIP models still assume a constant ligand concentration (as per the earliest Fe models) that is 1 nM for all models except BFM and PISCES1 who use 0.6 nM. Two exceptions in this regard are PISCES2 and TOPAZ. TOPAZ applies an empirical relationship to dissolved organic carbon (DOC) to derive ligand concentrations ($5x10^{-5}$ mol ligand per mol DOC). PISCES2 is the only FeMIP model to represent a dynamic ligand pool with explicit sources and sinks [*Völker and Tagliabue*, 2015] and a variable computation of the colloidal Fe fraction [*Liu and Millero*, 1999], modified to account for hydrothermal ligand supply [*Resing et al.*, 2015]. BLING

switches off Fe scavenging when oxygen drops below 1 mmol m^{-3} [*Galbraith et al.*, 2010] and both BLING and COBALT reduce the stability of Fe-ligand complexes in the presence of light [*Galbraith et al.*, 2010; *Stock et al.*, 2014]. Both the MITecco and MITigsm models cap DFe to a maximum value of 1.3 nM with any excess Fe being numerically deleted. Due to the noted flexibility in planktonic demands for Fe [*Sunda and Huntsman*, 1997; *Twining and Baines*, 2013], almost all FeMIP models have variable Fe/C ratios, with only MEDUSA1, MEDUSA2, MITecco and MITigsm retaining fixed Fe/C ratios. Recycling by zooplankton is variable in some FeMIP models and thus dependent on an assumed zooplankton Fe quota, except for BEC, BLING, MEDUSA1, MEDUSA2 and REcoM where there is a fixed rate of recycling. Lastly, all models include one particulate Fe pool, except PISCES1 and PISCES2 that consider 2 and BEC, which represents sinking implicitly (accounting for ballasting). Only COBALT invokes reduced regeneration efficiency relative to organic material that elongates the regeneration depth-scale beyond that that for sinking organic material [*Stock et al.*, 2014].

Finally, it is notable several models were only run for a few decades or centuries (BEC, BFM, COBALT, MEDUSA1, MEDUSA2, MITecco and MITigsm), a time comparable to the respective residence time of Fe in the model in some cases, making them potentially more sensitive to their initial conditions. This issue is discussed in more detail in Sec. 3.1.1.

3. Results

3.1 Inter-model differences in dissolved iron distributions and cycling

3.1.1 Iron fluxes and residence times

Beginning with an integrated view, there is substantial variability in the modeled Fe residence times across the FeMIP models with two broad groupings of a few years and a few hundred years (Table 2). Across the thirteen models, all include dust sources, ten include sediment sources, but only three include hydrothermal and riverine Fe sources, respectively (Table 2). Even for a given source, there is substantial inter-model difference in its strength. For example, dust fluxes of dissolved iron range from ~1 to >30 Gmol Fe yr^{-1} between models (Table 2, accounting for any inter-model variations in solubility and mineral fraction). These inter-model differences across all input fluxes result in a wide range of total iron inputs to the ocean (66.9±67.1 Gmol Fe yr^{-1}, Table 1). In contrast we find a surprising degree of agreement in the mean ocean iron concentration (0.58±0.14 nM, Table 2) from the models, with slightly greater inter-model differences in the total integrated inventory of Fe reflecting different model grid sizes (e.g. some models do not include the Arctic Ocean or the Mediterranean). Ultimately this results in a wide range of residence times of dissolved iron in the models (~5 to > 500 years, Table 2) that reflects different assumptions regarding the strength of the sources of DFe to the ocean, compensated by variable scavenging rates in order to reproduce the observed DFe concentration.

The derivation of the residence time for Fe from each model allows us to evaluate the impact of the shorter runs performed for some models. Taken at face value, even the relatively short runs performed by almost all the models (except BFM, MEDUSA1 and perhaps also MEDUSA2) are more than twice the residence time for Fe in that particular model. Nevertheless, it should be noted that many of these residence times for the global ocean are likely skewed towards lower values due to strong local sources that have a muted wider influence. For example, much of the interior Fe distribution in the PISCES1 model has been shown to be linked to a subducted preformed component [*Tagliabue et al.*, 2014b], suggesting that the deep ocean equilibration timescale in this model, at least, must be much longer than the 11 years of its

average residence time. This is likely to be the case for models that employ a formulation for the rate of DFe scavenging that depends on particulate fluxes, as biogenic fluxes in the ocean interior are considerably slower than near the surface where sedimentary and dust sources are dominant. Feedbacks will also exist between DFe inventory and biological fluxes, meaning that a ~1000 yr time-scale component to the overall equilibrium adjustment will exist that involves the redistribution of major nutrients globally. As such, this raises questions regarding the distributions of Fe in the ocean interior for models that are only run for a few decades, even if that is longer than the average residence time.

3.1.2 Statistical assessment of FeMIP models

In order to provide a general picture of variability amongst the models, we examine correlations between observed and simulated DFe at the same locations (Table 3). When viewed globally throughout the entire water column, correlations between observations and the models can be as high as 0.51, while some are even anti-correlated. The mean biases against observations are between -0.02 and -0.48 nM. In the 0-100m depth stratum, where Fe is likely be to playing a role in regulating phytoplankton growth rates, all but one of the model correlations fall between 0.33 and 0.48, implying no clear link between model complexity and strength of correlation. On the other hand, the mean biases range from -0.29 to 0.67 nM, which is suggests less overall agreement in the absolute DFe levels. The 100-500 m depth slice has the overall highest correlations, and all but three models reach their highest correlations in this depth range. In the abyssal layers only the three models that consider hydrothermal iron input (BEC, PISCES1 and PISCES2) show a reasonable correlation with observations (R=0.20 to 0.35, other models are < 0.15), highlighting the importance of this source in the deep ocean. However, the inclusion of hydrothermal iron input does not obviously lead to a significant improvement in the surface ocean. Similarly, including (or not) sedimentary Fe input does not seem closely linked to reproducing observations in the surface or intermediate layers. For example the two versions of MEDUSA with and without sedimentary iron input do not show much difference in their correlation coefficients. It is also important to note that we lack substantial coastal DFe datasets where sediments and/or river supply results in high DFe levels in a number of models (see Sec 3.1.3). Section 3.2 will more closely examine the different models using recent large-scale GEOTRACES sections as case studies in different ocean regions.

3.1.3 Inter-model differences in dissolved iron

To examine the inter-model differences in dissolved iron in more detail, we compare the model mean DFe over the 0-100m, 100-500m, 500-1000m and 2000-5000m depth slices, repeating the analysis for the boreal (30-90N), tropical (30N-30S) and austral latitudes (90S-30S). This enables us to group the models into 'high', 'moderate' and 'low' in terms of their DFe distribution, relative to the full model suite (Figure 1). Comparing Figure 1 with the statistical summary (Table 3) suggests that the inter-model trend in the average DFe concentration for the different depth slices does not always reflect good statistical agreement with the observations. However, it should be noted that while the inter-model trends in average DFe reflect full spatial and temporal averages, the statistics determined from observations only concern locations with available DFe observations (which is not spatially and temporally complete).

Beginning with the surface ocean (0-100m) that is heavily influenced by surface sources and biological uptake. MEDUSA1, MEDUSA2 and TOPAZ are consistently relatively high in iron for all three latitudinal zones, including the Fe limited Southern latitudes. BEC is also relatively

rich in Fe, but only in the northern and tropical latitudes. The lowest DFe concentrations in all three geographic zones are simulated by the BLING, COBALT and MITigsm models, with the remaining models intermediate throughout.

The relative tendencies between the different FeMIP models are generally conserved in the 100-500m and 500-1000m depth slices that are more heavily influenced by remineralisation processes. Notable departures from this general trend are PISCES2 displaying relatively higher DFe levels in both depth bins. While both BFM and REcoM become more DFe rich in the 500-1000m depth bin, TOPAZ stands out less as a high DFe model. In terms of hemispheric contrasts, BEC becomes lower in DFe in the Southern region; otherwise the inter-model trends are preserved.

In the deepest depth bin deep ocean sources such as hydrothermal vents, as well as sediments are important. Unsurprisingly, the models that include hydrothermal vent DFe sources (BEC, PISCES1 and PISCES2) show high DFe levels. In contrast, the high DFe levels for BFM, MITecco, MITigsm and REcoM cannot be ascribed to hydrothermal DFe input and may be related to initial conditions (e.g. for BFM) or deep ocean transport of high DFe levels. However, it is notable that BFM, MITecco, MITigsm and REcoM do not perform well statistically in this depth range (Table 3). The BLING and MEDUSA1 models simulate the lowest concentrations in this depth bin. For a large number of models (BLING, GENIE, MEDUSA1, MEDUSA2, TOPAZ), DFe concentrations decline in the 2000-5000m bin, relative to the 500-1000m bin.

3.1.4 Surface DFe distributions in the models

Due to its role as a limiting nutrient, we explore the simulated annual mean surface DFe concentrations from the FeMIP models in more detail (Figure 2, upper 50m average). Here we see that, as suggested by the range in the model biases (Table 3), there is a substantial degree of inter-model discord in the surface Fe distributions. Most models agree that the highest DFe concentrations are found underneath the Saharan dust plume in the tropical Atlantic, but others also emphasise dust supply into the Arabian Sea and enhanced DFe along the continental margins. A large number of the models suggest the lowest DFe concentrations are found across the Pacific Ocean. Exceptions are GENIE and MEDUSA1, who have much higher DFe concentrations therein and BEC, MEDUSA2 and TOPAZ, who restrict low DFe to the south Pacific only. The sub-Arctic Pacific is much more DFe deplete in BFM, MITecco and MITigsm relative to the other FeMIP models. When the seasonality in DFe (presented as the maximum minus minimum DFe concentration over the year, Figure 3) is compared, strong inter-model differences also emerge. For example, some models show remarkably little seasonality (BFM, GENIE, MEDUSA1, MEDUSA2 and MITigsm), whereas others have large seasonal cycles over wide areas (>0.5nM, BEC, MITecco, PISCES1, PISCES2 and TOPAZ). This illustrates where high annual mean concentrations in these regions are masking strong seasonal minima. For this reason it is not straightforward to compare the models against observed Fe that might have been collected during different seasons. At this stage, incomplete sampling over the seasonal cycle is prevalent for virtually all locations with DFe measurements [*Tagliabue et al.*, 2012], which precludes the mapping of DFe seasonality from observations. Table 3 is therefore more suited for a statistical assessment of the surface DFe for a given model against all available observations (where seasonal variations are accounted for by comparing model and data DFe at identical longitudes, latitudes, depths and months).

3.2 Comparison to recent GEOTRACES ocean sections

To more closely examine how the different DFe models represent the observed distribution of DFe we focus on a range of recent GEOTRACES sections. As described above (Sec 2.2) each model is extracted at the exact location of the sampling locations, with the observations regridded onto the same vertical grid. We refer the readers to the below cited papers for a more complete discussion of each observational section and additional interpretation. In this assessment we emphasise the key features observed on each section and how different models are able to reproduce them. Because of this goal and because a given model may do a good job of reproducing one feature, but not another, we did not perform statistical assessments of the individual models for each section.

3.2.1 West Atlantic

The GA-02 West Atlantic meridional section provides unprecedented coverage of DFe concentrations along the Atlantic Ocean, as well as insights into different mechanisms that control the cycling, regeneration and supply of DFe [*Rijkenberg et al.*, 2014]. The key features of this section are (i) low surface DFe in both the northern and southern end member surface waters, (ii) a surface DFe enrichment around 20^{o}N in the tropics and associated with a subsurface DFe minima, (iii) a strong DFe regeneration maxima at 5-10^{o}N centered around 500-1000m, (iv) a hydrothermal signal at around 5^{o}S and between 2000-3000m depth and (v) a hotspot of DFe that is present over much of the water column associated with the confluence of the Brazil and Falklands current at around 35-40^{o}S.

Model representation of key features (Figure 4): (i) Almost all models capture low DFe in the Southern end member surface waters, except MEDUSA1 and MEDUSA2 and perhaps also REcoM and TOPAZ. However it is only in BFM and COBALT, and to a lesser degree BEC, BLING, MITigsm, PISCES1 and PISCES2 that reproduce the observed low DFE concentrations associated with the northern endmember surface waters. (ii) A surface DFe enrichment (presumably from dust) around 20^{o}N is clearly present in BEC, MEDUSA1, MEDUSA2, MITecco, MITigsm, PISCES1, PISCES2, REcoM and TOPAZ, but is less apparent in other models (BFM, BLING, COBALT, GENIE). Nevertheless, in MEDUSA2, REcoM and TOPAZ the influence of surface dust deposition appears to be much greater than is observed. Only PISCES1, PISCES2 and COBALT, show the observed subsurface minima in DFe below the dust signal. (iii) With respect to the strong DFe regeneration maxima at 5-10^{o}N centered around 500-1000m, COBALT displays a regeneration maximum at around the right depth level, while in BEC high concentrations appear to be smeared from surface to the sea floor. In all other models the regeneration signal in DFe is generally too small or absent and where it is present (e.g. BFM, BLING, GENIE) it is generally too shallow in the water column. (iv) Concerning the hydrothermal signal at around 5^{o}S and between 2000-3000m depth, of the three models that include hydrothermal DFe input, only PISCES2, with a greater longevity of hydrothermal Fe [*Resing et al.*, 2015], shows a hint of DFe enrichment in the right location. MEDUSA2 underestimates DFe in the ocean interior along the entire Atlantic section. (v) No models capture the elevated DFe over almost the entire water column around 35-40^{o}S. In the observations, this is ascribed to the offshore export of Brazilian shelf waters or DFe input from the dissolution of particulate Fe associated with the Rio de la Plata river [*Rijkenberg et al.*, 2014].

3.2.2 Subtropical North Atlantic

The GA-03 North Atlantic zonal section crossed the subtropical North Atlantic between Cape Verde and Woods Hole (USA) via Bermuda. Key signals in the dataset [*Hatta et al.*, 2014] are (i) strong enhancements in DFe associated with DFe regeneration and also coastal input along

the eastern and western margins, (ii) a surface enrichment along with a subsurface minimum in DFe and (iii) a strong hydrothermal anomaly over the mid Atlantic ridge.

Model representation of key features (Figure 5): (i) Enhanced DFe in the subsurface along the margins is represented to different degrees by the FeMIP models. BLING, COBALT, MITecco and PISCES1 have hints of subsurface maxima in DFe along the eastern margin. It is encouraging that the addition of ligand production during remineralisation in PISCES2 clearly improves the intensity of the remineralised DFe signal. However, none of these models have a broad homogenous signal (down to > 2000m) of elevated DFe that is observed on the eastern margin, except perhaps BEC, which has a strong subsurface maximum that spreads over all depth levels. (ii) The subsurface minima in DFe underlying a surface (presumably dust) enrichment is captured clearly by COBALT, PISCES1 and PISCES2 and slightly less clearly by BEC, BFM and BLING. (iii) A hydrothermal anomaly is present in PISCES1, but closer in magnitude to the observations in PISCES2, while BEC also displays a strong hydrothermal signal. COBALT displays a sediment signal at depth that is not reproduced by the observations. It also notable that many of the models present an 'inverted' DFe profile, with decreasing DFe concentrations towards the ocean interior (GENIE, MEDUSA1, MEDUSA2, REcoM and TOPAZ), which could be indicative of too great a residence time for DFe at the ocean surface. Also, BLING, COBALT, MEDUSA2 and TOPAZ seem to be systematically too low in terms of their interior ocean DFe levels across this section.

3.2.3 Subtropical South Atlantic

The CoFeMUG section traversed the south Atlantic between Namibia and Brazil and had the following notable signatures [*Noble et al.*, 2012; *Saito et al.*, 2013]: (i) a remineralisation signal and/or sediment input on the eastern margin, (ii) low overall surface concentrations and (iii) a strong hydrothermal signal at depth.

Model representation of key features (Figure 6): (i) Interestingly, more models are able to simulate a remineralisation signal on the eastern side of the basin (COBALT, MEDUSA1, MEDUSA2, MITecco, PISCES1, PISCES2, REcoM and TOPAZ) for this section than for the GA03 section. Although for some models this feature is too weak or spread over too many depth levels. (ii) All models, except MEDUSA1, MEDUSA2, REcoM and TOPAZ, are able to reproduce the overall low DFe conditions in the surface waters. (iii) BEC and PISCES1 represent a DFe anomaly over the ridge as observed, but this is underestimated. PISCES2 represents a stronger hydrothermal signal, but it appears to spread too far off-axis relative to that observed. Again, COBALT displays a strong sediment signal in the deep ocean that is not observed. BFM, BLING, MEDUSA2 and to some extent TOPAZ underestimate interior ocean DFe levels.

3.2.4 Southern Tropical Pacific

The GP-16 cruise ran from Ecuador to Tahiti [*Resing et al.*, 2015] and displays the following key features: (i) DFe enrichment along the eastern margin over almost the entire water column, (ii) low surface concentrations and (iii) a remarkable hydrothermal plume propagating westward for > 4000km from the East Pacific Rise to at least 150^{o}W.

Model representation of key features: (Figure 7), (i) BEC, COBALT, PISCES2 and TOPAZ are the only models able to produce the broad signal of elevated DFe throughout the entire water column on the eastern margin. BLING, MEDUSA1, MEDUSA2 and REcoM display an enrichment in DFe but this remains more tightly localised than observed. (ii) All models

capture the low DFe levels typical of Pacific surface waters, but for some models (BFM, BLING, COBALT, GENIE, MEDUSA2 and TOPAZ), low DFe is also too prevalent in the ocean interior. (iii) BEC and PISCES1 capture a local hydrothermal signal above the East Pacific Rise, but only PISCES2 goes any way towards reproducing the degree of off axis transport. As seen previously, MITigsm and COBALT show DFe increases near the sea floor, but these are more widespread than seen in the observations. As noted previously, BFM, BLING COBALT, MEDUSA2 and TOPAZ show too little DFe in the ocean interior (<0.3nM), relative to the observations (>0.6 nM away from the hydrothermal plume).

3.2.5 Southern Ocean – Atlantic Sector

Both the GIPY-4 and GIPY-5 cruises ran from Cape Town (South Africa) to the Antarctic continent along the so-called 'GoodHope' line during the International Polar Year [*Chever et al.*, 2010; *Klunder et al.*, 2011]. These cruises sampled at different resolutions north and south of the Polar Front and have been blended to form one section. Notable features in this dataset include (i) low but non zero concentrations at the surface that propagate into the subsurface, (ii) a strong remineralisation signal at around 500m near 60S and (iii) a strongly local hydrothermal signal over the Bouvet region ridge crest at around 54°S and more widespread elevated DFe in the abyssal ocean north of the ridge (i.e. between ~54°S and the northern end of the transect.

Model representation of key features: (Figure 8), (i) Most models display low overall DFe concentrations at the surface. GENIE, MEDUSA1, MEDUSA2 and to a lesser degree REcoM and TOPAZ over estimate surface DFe concentrations. But even the models that have low surface DFe show rapid increases with depth, indicating that the ferricline is too shallow in all models. (ii) No FeMIP model captures the remineralisation signal seen in the subsurface just south of the Polar Front. (iii) Despite including a hydrothermal source, BEC is unable to represent the local hydrothermal enrichment. While PISCES1 represents a slight hydrothermal anomaly that appears to be from an adjacent source, the longer lifetime of hydrothermal Fe in PISCES2 leads to the anomaly being too widespread in the abyssal ocean. On the other hand, both BEC and PISCES2 do show elevated DFe in the abyssal ocean north of the main ridge at 54°S that compares well with the data. COBALT, MITecco and MITigsm again show a sediment signal in DFe at depth, while COBALT and TOPAZ show very high values near the Antarctic coast. None of these features are observed in the dataset. The BFM stands out from the other models with the large underestimation of DFe in the Southern Ocean interior as already seen for the GA-02 section.

4. Discussion

4.1 Examining inter-model differences in Fe distributions relative to other nutrients

In short, we find a wide range of simulated DFe distributions from current global ocean biogeochemical models that reflects an apparent lack of inter-model agreement in the processes that control the oceanic distribution of DFe. When assessed against the best DFe datasets, most models perform modestly both quantitatively in terms of magnitudes and patterns, and qualitatively in representing the inferred mechanisms. This has important implications for how models are used to understand biogeochemical cycles [*Galbraith et al.*, 2010; *Moore et al.*, 2002; *Tagliabue et al.*, 2014a], planktonic diversity and resource competition [*Dutkiewicz et al.*, 2012; *Ward et al.*, 2013], as well as the ocean response to fluctuations in the environment in general [*Bopp et al.*, 2013; *Dutkiewicz et al.*, 2013; *Tagliabue et al.*, 2009]. It is noteworthy that this inter-model disagreement appears to be

511 solely driven by the particular way in which different models represent the Fe cycle. If we
512 examine the models in terms of macronutrients (nitrate and phosphate) then, taking the long
513 meridional GA02 section as example, we see a much stronger inter-model and model-data
514 agreement (Figures 9 and 10). Although inter-model differences due to specific physical
515 models are visible in the Atlantic water mass structure, the mechanisms driving the N and P
516 cycles are similar.
517
518 We further contextualise the inter-model Fe differences by examining how they represent the
519 relative inventories of Fe and NO_3 in the ocean interior by plotting the Fe* tracer (Fe –
520 $NO_3 * r_{Fe/N}$). Defining $r_{Fe/N}$ in the same way as for the GA02 section [*Rijkenberg et al.*, 2014]
521 (based on the observed Fe:apparent oxygen utilisation relationship, which results in a Fe/N
522 ratio of 0.47 mmol/mol) and using PO_4 (and a NO_3/PO_4 ratio of 16/1) for GENIE and BLING,
523 which do not simulate NO_3, allows us to examine DFe concentrations relative to NO_3, (Figure
524 11). The data shows relatively replete waters originating from the northern hemisphere
525 linked to North Atlantic Deep Water (NADW), which becomes flanked above and below by
526 relatively Fe poor water from the southern hemisphere linked to Antarctic Intermediate
527 Water (AAIW) and Antarctic Bottom Water (AABW). There is also a zone of relatively
528 depleted Fe in the subsurface overlying the NADW signal in the northern hemisphere likely
529 linked to northern subtropical mode water. In these sections we can see that NADW is
530 relatively impoverished in DFe in MEDUSA1, MEDUSA2 and TOPAZ, despite these models
531 generally overestimating surface DFe. This may indicate an overly short lifetime for Fe away
532 from the surface and subsequent lack of permanence in the NADW signal. Looking at
533 southern sourced waters, all models except BFM perform well (notwithstanding the northern
534 sourced water biases). Obviously, this comparison should only be taken as indicative since
535 different models are underpinned by different relationships between NO_3 and Fe and the
536 actual planktonic Fe:N ratio can vary from the value chosen in the *Rijkenberg et al.* [2014]
537 study [*Twining and Baines*, 2013]. Nevertheless, it does provide an additional means to assess
538 the relative transport of Fe and NO_3 through the ocean interior.
539
540 **4.2 Identifying the key processes at different depth strata**
541
542 One important inter-model difference that clearly impacts the agreement with observations
543 and the role of Fe on biota is the strong surface enrichments evident in some models
544 (MEDUSA1, MEDUSA2, REcoM and TOPAZ). In the observations, any Fe enrichments due to
545 dust deposition are far more localised and apparently short lived in space (e.g. Figures 4 and
546 5). For the models surface overestimation of iron implies either too large an iron source or
547 that the residence time for Fe at the surface is too long. The latter possibility highlights the
548 importance of how models treat the scavenging process and could also be linked to constant
549 Fe/C ratios that do not permit 'luxury uptake' of Fe at high DFe concentrations (specifically
550 MEDUSA1, MEDUSA2, MITecco and MITigsm). MEDUSA1, MEDUSA2 and REcoM are three of
551 the four models with the longest residence times (decades to centuries, Table 2), relative to
552 the other FeMIP models, and produce high surface enrichment despite having some of the
553 lowest dust inputs (Table 2). For MEDUSA1 and MEDUSA2 the first order fixed scavenging
554 rate may be too low or have not enough variability to remove Fe rapidly when concentrations
555 are high. The constant Fe/C ratios used in these two models may also contribute to this
556 anomalous feature. In REcoM, Fe/C ratios are variable and the scavenging is second order, but
557 may simply be too low. DFe in TOPAZ has one of the shortest residence times (~8 years, Table
558 2), which implies that the surface accumulation of DFe may instead be linked to relatively
559 large sources or the variable ligand concentration. Since the ligand concentration in TOPAZ
560 depends on DOC, which typically decays from surface to deep, there may be too much DFe
561 stabilisation occurring in the surface ocean.

At intermediate depths, the inclusion of a prognostic Fe binding ligand pool with a particle degradation source [*Völker and Tagliabue*, 2015] clearly improves the reproduction of subsurface maxima in DFe associated with remineralisation (compare PISCES2 with PISCES1) for many of the transects. Other models (COBALT and to a lesser degree BEC and BLING) are able to reproduce these features but evidently do so for different reasons. These may be related to the implicit formulation of particle flux (BEC) that ignores lateral transport of particulate Fe or the shutdown of Fe scavenging in low oxygen conditions (BLING). It is interesting that there appears to be two groups of subsurface DFe maxima seen in the observations. Sometimes these features are tightly constrained to a small depth stratum (e.g. equatorial ocean for GA-02, western margin on GA-03 and eastern margin on CoFeMUG), while in other locations the DFe enrichments span almost the entire water column (eastern margins on GA-03 and GP-16). Most models represent one or the other. For example, subsurface maxima are always tightly bounded in depth for some models (e.g. COBALT and PISCES2) or spread over depth in others (BEC) with no regional variations. Future work should explore the potential mechanisms involved, which might be linked to subsurface dissolution of dust, nutrient trapping or impacts of low oxygen. Emerging Fe isotope work highlights the potential for non-reductive Fe release from margins [*Conway and John*, 2014; *Homoky et al.*, 2013] in addition to the role of reducing sediments represented in models.

In the ocean interior the best models (in terms of their linear correlation coefficients) are those that include hydrothermal input (Table 3). While including such a source is clearly important, it is possible that this is overemphasised in the correlations at the expense of other deep ocean structure that is evident in many of the sections. For example, many of the ocean sections do not show any 'watermass' related structure for DFe that is seen in macronutrients (e.g. Figures 9 and 10). Although adding a hydrothermal ligand seems to improve the ability of PISCES2 to reproduce the GP-16 data (Figure 7) and perhaps also the GA-02 hydrothermal signal (Figure 4), it results in too widespread a hydrothermal anomaly in the Southern Ocean (Figure 8) indicating too long a lifetime for this pool and the need for further refinement of the processes governing hydrothermal Fe input [*Tagliabue*, 2014].

4.2 Inter-Model differences in DFe inputs and cycling: the importance of scavenging

It is notable that there is a great deal of variability in both the total Fe input flux (66.9±67.1 Gmol Fe yr^{-1}) and the strength of a given source across the models, yet the mean ocean DFe is strikingly similar (0.58±0.14 nM). To a large extent, this agreement reflects the calibration of scavenging rates and the concentration of organic ligands to obtain global average iron concentrations in agreement with observations. While this relative homogeneity in modeled mean DFe would be consistent with an earlier view of the oceanic Fe inventory [*Johnson et al.*, 1997], if anything, the emerging oceanic sections of DFe as part of the GEOTRACES programme have highlighted an unexpected variability in DFe distributions in the ocean interior [*Mawji et al.*, 2015]. This is in stark contrast to the other main limiting nutrients, which more closely reflect large-scale ocean circulation patterns and watermass related features (e.g. Figures 9 and 10). Thus the apparent small differences in the mean ocean DFe between models more likely arises from a modeling community that reflects an earlier parsimonious view of the system. The relative constancy in the mean ocean DFe concentrations in the models may reflect homogenous ligand concentrations of either 0.6 or 1.0 nM, but we note that even models with varying ligand concentrations (PISCES2 and TOPAZ) show too much interior ocean uniformity.

In contrast to the mean DFe, there is a substantial degree of inter-model disagreement in the strength of different sources. For instance, BFM, BLING, GENIE, MEDUSA1, MEDUSA2, MITecco, MITigsm and REcoM all have atmospheric input fluxes of < 5 Gmol Fe yr^{-1}, whereas as in BEC, COBALT, PISCES1, PISCES2 and TOPAZ dust supply is much higher (> 20 Gmol Fe yr^{-1}). Yet this does not drive a similar trend in mean ocean DFe (with MITecco, MITigsm and REcoM showing amongst the highest DFe concentrations, Table 2). We note that these represent the total DFe flux from dust, accounting for model specific Fe mineralogy and solubility. Equally, for those models that include sedimentary Fe input, this flux term can range from very small (e.g. < 5 Gmol Fe yr^{-1} in MEDUSA2 or REcoM) to very large (> 70 Gmol Fe yr^{-1} in BEC, COBALT, MITecco, MITigsm and TOPAZ). Again this does not map onto mean DFe trends. We note that the closer agreement for hydrothermal Fe input is more likely to reflect the fact that only two models actually include this term, rather than greater confidence regarding the actual flux. Overall, the total input of DFe does not explain the inter-model variations found in mean DFe (R^2=0.06). This implies that there must be a great deal of variability in how each model treats the scavenging of Fe in order to ultimately arrive at a relatively similar mean ocean DFe concentration.

Most early Fe models that explicitly computed free Fe and sought to represent its scavenging by sinking particles, treated the scavenging rate constant as a tunable parameter [*Archer and Johnson*, 2000; *Johnson et al.*, 1997; *Parekh et al.*, 2004; *Watson et al.*, 2000]. This was viable in these relatively simple box models against few observations, but is a less straightforward solution for the multi tracer/process 3D biogeochemical models used presently where scavenging itself maybe a function of other model parameters (e.g. particle concentrations) and hence can vary considerably in space. Despite the long acknowledged influence of the particle concentration on the scavenging rate [*Honeyman et al.*, 1988], a subset of the FeMIP models persist with a globally uniform scavenging rate (Table 1). However, even for those models that have implemented a second order scavenging rate, there is a question of how this should operate. For example, should the model rely only on organic carbon or also include biogenic silica and calcium carbonate? Non-biogenic particles, such as dust, as well as Fe and manganese oxides, may also be important as Fe scavengers [*Hayes et al.*, 2015; *Wagener et al.*, 2008; *Ye et al.*, 2011]. There is also the important question of the specific affinity for free Fe for these various carrier phases. Once Fe is scavenged onto particles, desorption of Fe will be important in resupplying the DFe pool. Some models consider constant desorption rates [*Moore and Braucher*, 2008], while others explicitly account for disaggregation dynamics and the impact of bacterial activity [*Aumont et al.*, 2015]. Finally, there is the question of regional and temporal variability in colloidal dynamics. Only some FeMIP models attempt to account for this process (Table 1), yet given the apparent importance of colloidal Fe within the DFe fraction [*Boye et al.*, 2010; *Fitzsimmons and Boyle*, 2014; *Wu et al.*, 2001], colloidal pumping losses might be as large as those from the scavenging of free Fe. Some progress may be made by exploiting the legacy from the field of Thorium (Th) cycling, for which a number of different theories have been developed to describe its scavenging, including colloidal components [*Anderson*, 2003; *Burd et al.*, 2000; *Lam and Marchal*, 2015; *Marchal and Lam*, 2012; *Savoye et al.*, 2006]. With an expanding database of paired Fe and Th observations, including the particulate phase, as part of GEOTRACES [*Mawji et al.*, 2015] it may be possible to refine this crucial component of the Fe cycle in the coming years.

4.3 Impact of Fe on wider biogeochemical cycles: the importance of biological Fe cycling

The biological cycling of DFe in a given model will dictate the net influence of a model's DFe cycling on wider biogeochemical cycling and air-sea CO_2 exchange. In that regard, the large oceanic sections, focused process studies and laboratory experiments all provide essential

and complementary information. For example, early laboratory studies demonstrated a large degree of flexibility in the phytoplankton Fe/C ratios as a function of DFe levels and cell size, as well as enhanced Fe/C ratios at lower light levels [*Sunda and Huntsman*, 1997]. Similar ranges in Fe/C ratios are also seen in single cell analyses of phytoplankton from the ocean [*Twining and Baines*, 2013]. The enhanced Fe/C ratio seen at low light is thought to reflect so-called 'biodilution', where Fe uptake continues when phytoplankton carbon fixation is light limited, and/or a greater absolute demand for Fe at low light [*Sunda and Huntsman*, 1997; *Sunda and Huntsman*, 1998]. Almost all FeMIP models permit flexibility in the Fe/C ratio of phytoplankton (Table 1), with those that consider Fe uptake independent of C fixation able to account for any biodilution and the BLING model considers a direct impact of Fe on photosynthesis. Emerging recent work has suggested that there are important inter-specific differences in how phytoplankton Fe demands respond to light [*Strzepek et al.*, 2012]. In their laboratory study, *Strzepek et al.* [2012] found that while temperate diatom species indeed showed elevated Fe/C ratios at low light, the opposite was true for Antarctic diatom species. This raises questions about how models that generally do not consider different phytoplankton species (but rather represent broader 'functional types') can account for these potentially important regional distinctions in how environmental variations impact biological Fe cycling.

Detailed process studies, mostly from the Southern Ocean, have sought to quantify Fe cycling at the ecosystem level. In doing so, the importance of regenerated Fe in the fuelling of biological productivity via the so-called 'ferrous wheel' has emerged as potentially important [*Bowie et al.*, 2009; *Bowie et al.*, 2015; *Boyd et al.*, 2012; *Boyd et al.*, 2005; *Sarthou et al.*, 2008; *Strzepek et al.*, 2005]. This has been demonstrated via the development of the '*f*e-ratio', which represents the proportion of Fe uptake from 'new' Fe sources. It has been determined for sites across the Southern Ocean by assembling Fe budgets that combine measurements of Fe pools and fluxes alongside laboratory estimates. The *f*e-ratio is generally around 0.1 (i.e. strongly reliant on recycled Fe) in the low productivity regions of the Southern Ocean [*Bowie et al.*, 2009; *Boyd et al.*, 2005] and reaches around 0.5 and greater (i.e. less reliant on recycled Fe) in the naturally fertilised Kerguelen Island phytoplankton bloom [*Bowie et al.*, 2015; *Sarthou et al.*, 2008]. Langrangian process studies have demonstrated a strong seasonal decline in the *f*e-ratio as the spring phytoplankton bloom declines [*Boyd et al.*, 2012], which are consistent with low rates of Fe input during summer [*Tagliabue et al.*, 2014c]. In agreement, direct measurements of Fe fluxes between various components of the food web have highlighted that only regenerative fluxes can support the measured Fe demand [*Boyd et al.*, 2012; *Strzepek et al.*, 2005; *Tagliabue et al.*, 2014c].

The sensitivity of a given model's biological productivity to new or regenerated forms of Fe is crucial, as this will underpin its sensitivity to change. At present we do not know if the FeMIP models place the correct emphases on new and recycled Fe in different ocean regions. Many models rely on fixed rates of Fe regenerated by zooplankton and the remineralisation of organic material, while others allow this to vary (Table 1). A key parameter in driving the turnover of Fe by the zooplankton and bacterial communities in such models is an estimate of the heterotroph demand for Fe, which is then balanced against the Fe/C provided as nutrition. New measurements of stocks and turnover of Fe from specific ocean regions are also beginning to emerge [*Boyd et al.*, 2015], which will be invaluable in assessing the magnitude and variability of the modelled rates.

5. Future Work

A weakness of the current intercomparison is that we did not truly intercompare the Fe models, but instead compared the models' coupled physical-biogeochemical framework (including Fe). This was necessary to retain as broad a suite of models as possible for this first intercomparison. In future work, it would be useful to intercompare different Fe models within the same physical model framework (e.g. as possible in the NEMO or MITgcm modelling frameworks). Additionally, a set of planned model perturbations could be performed where each individual model is subjected to a modification to its Fe supply (either as a direct fertilisation event or by an alteration to one of the input fields). Much could be learned from the way the Fe cycle responds to such perturbations across the different models.

Reducing uncertainty in the input fluxes of Fe is clearly important, but has proved difficult to achieve over recent years (even for long recognised Fe sources such as dust). Some progress could be made by implementing 'source specific' tracers (such as aluminium or manganese) alongside Fe to constrain individual sources. Constraining scavenging rates has emerged as a key priority and parallel simulation of Th may help constrain rates of Fe loss and the particle pools. Moreover, many of the models used specifically for ecological questions are only run for a few decades, leading to a greater sensitivity to initial conditions. A priority for such 'resource intensive' models would be the availability of input fields based on data climatologies (such as those available for macronutrients as part of the World Ocean Atlas datasets) or consensus distributions that may emerge from improved models.

As described in Sec. 4.3 an assessment of the different biological Fe models is also a priority, as this will underpin the carbon cycle response and has not been compared against the paradigms recently emerging from experimental work. A follow-up Phase of FeMIP could include a closer comparison of the models against the detailed process study measurements made (for example) as part of the FeCycle set of experiments [*Boyd et al.*, 2012; *Boyd et al.*, 2005]. A range of the Fe models could be set up in a one dimensional lagrangian framework and forced by observed physics to be compared rigorously against the measured Fe stocks and cycling rates.

6. Conclusions

We have compared the projected DFe distributions from thirteen global ocean biogeochemistry models against each other and with available datasets. Newly-available full depth sections of DFe collected from different oceanic regions as part of the GEOTRACES programme have greatly facilitated this task. All models do relatively poorly in reproducing a global DFe dataset of around 20,000 observations, which highlights the need for greater understanding of how the ocean Fe cycle functions and how Fe should be represented in global ocean models. We find a large degree of inter-model variability in the input fluxes of DFe, which leads to great variability in the modeled residence times. The stronger inter-model agreement in the mean ocean DFe most likely reflects earlier views of constant deep ocean DFe levels maintained by a homogenous ligand pool and requires calibration via poorly constrained scavenging rates. The way different models treat DFe scavenging has emerged as a key uncertainty that would benefit from stronger observational constraints. More detailed inter-model tests, particularly linked to process study data, are needed to assess the models' biological components.

In closing, we re-emphasise the importance of the iron cycle in global ocean biogeochemistry models, given its role, alongside NO_3, as one of the two most important limiting nutrients. Although the models analysed here struggle to capture the detailed distribution of this highly dynamic element, it is very likely that biogeochemical models that include an iron cycle can

produce a more realistic simulation than models that do not. Improving the quantitative understanding of iron cycling should be a major priority for ocean biogeochemistry research.

7. Acknowledgements

We thank everyone that has been involved in the development of the FeMIP ocean models over many years. A.T especially thanks Eric Achterberg, Andrew Bowie, Maarten Klunder, Joe Resing, Micha Rijkenberg, Mak Saito, Christian Schlosser and Peter Sedwick for sharing DFe datasets ahead of their publication and Edward Mawji, Reiner Schlitzer, Elena Masferrer-Dodas and the wider GEOTRACES community for their efforts in producing the Intermediate Data Product (2014) that facilitated this inter comparison. PISCES1 and PISCES2 simulations made use of the N8 HPC facilities, funded by the N8 consortium and EPSRC grant EP/K000225/1. MV acknowledges the BFM system team (http://bfm-community.eu) for the public availability of the BFM model. We thank Bob Anderson and Ric Williams for comments on the draft manuscript and those of two anonymous reviewers that improved the final version

8. References

Anderson, R. F. (2003), Chemical Tracers of Particle Transport, *Treatise on Geochemistry*, 247-273. doi:10.1016/b0-08-043751-6/06111-9

Archer, D. E., and K. Johnson (2000), A model of the iron cycle in the ocean, *Global Biogeochemical Cycles*, *14*(1), 269-279. doi:10.1029/1999gb900053

Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray (2003), An ecosystem model of the global ocean including Fe, Si, P colimitations, *Global Biogeochemical Cycles*, *17*(2). doi:10.1029/2001gb001745

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen (2015), PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, *Geoscientific Model Development*, *8*(8), 2465-2513. doi:10.5194/gmd-8-2465-2015

Barbeau, K., J. W. Moffett, D. A. Caron, P. L. Croot, and D. L. Erdner (1996), Role of protozoan grazing in relieving iron limitation of phytoplankton, *Nature*, *380*(6569), 61-64. doi:10.1038/380061a0

Bopp, L., et al. (2013), Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, *Biogeosciences*, *10*(10), 6225-6245. doi:10.5194/bg-10-6225-2013

Bowie, A. R., D. Lannuzel, T. A. Remenyi, T. Wagener, P. J. Lam, P. W. Boyd, C. Guieu, A. T. Townsend, and T. W. Trull (2009), Biogeochemical iron budgets of the Southern Ocean south of Australia: Decoupling of iron and nutrient cycles in the subantarctic zone by the summertime supply, *Global Biogeochemical Cycles*, *23*(4). doi:10.1029/2009gb003500

Bowie, A. R., et al. (2015), Iron budgets for three distinct biogeochemical sites around the Kerguelen Archipelago (Southern Ocean) during the natural fertilisation study, KEOPS-2, *Biogeosciences*, *12*(14), 4421-4445. doi:10.5194/bg-12-4421-2015

Boyd, P. W., and M. J. Ellwood (2010), The biogeochemical cycle of iron in the ocean, *Nature Geoscience*, *3*(10), 675-682. doi:10.1038/ngeo964

Boyd, P. W., R. F. Strzepek, M. J. Ellwood, D. A. Hutchins, S. D. Nodder, B. S. Twining, and S. W. Wilhelm (2015), Why are biotic iron pools uniform across high- and low-iron pelagic ecosystems?, *Global Biogeochemical Cycles*, *29*(7), 1028-1043. doi:10.1002/2014gb005014

Boyd, P. W., et al. (2012), Microbial control of diatom bloom dynamics in the open ocean, *Geophysical Research Letters*, *39*(18). doi:10.1029/2012gl053448

Boyd, P. W., et al. (2005), FeCycle: Attempting an iron biogeochemical budget from a mesoscale SF6 tracer experiment in unperturbed low iron waters, *Global Biogeochemical Cycles*, *19*(4). doi:10.1029/2005gb002494

Boye, M., J. Nishioka, P. Croot, P. Laan, K. R. Timmermans, V. H. Strass, S. Takeda, and H. J. W. de Baar (2010), Significant portion of dissolved organic Fe complexes in fact is Fe colloids, *Marine Chemistry*, *122*(1-4), 20-27. doi:10.1016/j.marchem.2010.09.001

Bruland, K. W., R. Middag, and M. C. Lohan (2014), Controls of Trace Metals in Seawater, *Treatise on Geochemistry, 2nd edition*, 19-51. doi:10.1016/b978-0-08-095975-7.00602-1

Burd, A. B., S. B. Moran, and G. A. Jackson (2000), A coupled adsorption–aggregation model of the POC/ ratio of marine particles, *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(1), 103-120. doi:10.1016/s0967-0637(99)00047-3

Chever, F., E. Bucciarelli, G. Sarthou, S. Speich, M. Arhan, P. Penven, and A. Tagliabue (2010), Physical speciation of iron in the Atlantic sector of the Southern Ocean along a transect from the subtropical domain to the Weddell Sea Gyre, *J Geophys Res-Oceans*, *115*(C10). doi:10.1029/2009jc005880

Conway, T. M., and S. G. John (2014), Quantification of dissolved iron sources to the North Atlantic Ocean, *Nature*, *511*(7508), 212-215. doi:10.1038/nature13482

Dunne, J. P., et al. (2013), GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics*, *Journal of Climate*, *26*(7), 2247-2267. doi:10.1175/jcli-d-12-00150.1

Dutkiewicz, S., J. R. Scott, and M. J. Follows (2013), Winners and losers: Ecological and biogeochemical changes in a warming ocean, *Global Biogeochemical Cycles*, *27*(2), 463-477. doi:10.1002/gbc.20042

Dutkiewicz, S., B. A. Ward, F. Monteiro, and M. J. Follows (2012), Interconnection of nitrogen fixers and iron in the Pacific Ocean: Theory and numerical simulations, *Global Biogeochemical Cycles*, *26*(1). doi:10.1029/2011gb004039

Dutkiewicz, S., B. A. Ward, J. R. Scott, and M. J. Follows (2014), Understanding predicted shifts in diazotroph biogeography using resource competition theory, *Biogeosciences*, *11*(19), 5445-5461. doi:10.5194/bg-11-5445-2014

Dutkiewicz, S., A. E. Hickman, O. Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows (2015), Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, *Biogeosciences*, *12*(14), 4447-4481. doi:10.5194/bg-12-4447-2015

Elrod, V. A., W. M. Berelson, K. H. Coale, and K. S. Johnson (2004), The flux of iron from continental shelf sediments: A missing source for global budgets, *Geophysical Research Letters*, *31*(12). doi:10.1029/2004gl020216

Fitzsimmons, J. N., and E. A. Boyle (2014), Both soluble and colloidal iron phases control dissolved iron variability in the tropical North Atlantic Ocean, *Geochimica et Cosmochimica Acta*, *125*, 539-550. doi:10.1016/j.gca.2013.10.032

Frew, R. D., D. A. Hutchins, S. Nodder, S. Sanudo-Wilhelmy, A. Tovar-Sanchez, K. Leblanc, C. E. Hare, and P. W. Boyd (2006), Particulate iron dynamics during FeCycle in subantarctic waters southeast of New Zealand, *Global Biogeochemical Cycles*, *20*(1). doi:10.1029/2005gb002558

Galbraith, E. D., A. Gnanadesikan, J. P. Dunne, and M. R. Hiscock (2010), Regional impacts of iron-light colimitation in a global biogeochemical model, *Biogeosciences*, *7*(3), 1043-1064. doi:10.5194/bg-7-1043-2010

Gerringa, L. J. A., A. C. Alderkamp, P. Laan, C. E. Thuroczy, H. J. W. De Baar, M. M. Mills, G. L. van Dijken, H. van Haren, and K. R. Arrigo (2012), Iron from melting glaciers fuels the phytoplankton blooms in Amundsen Sea (Southern Ocean): Iron biogeochemistry, *Deep-Sea Res Pt Ii*, *71-76*, 16-31. doi:Doi 10.1016/J.Dsr2.2012.03.007

Gledhill, M., and K. N. Buck (2012), The organic complexation of iron in the marine environment: a review, *Frontiers in microbiology*, *3*, 69. doi:10.3389/fmicb.2012.00069

Hatta, M., C. I. Measures, J. Wu, S. Roshan, J. N. Fitzsimmons, P. Sedwick, and P. Morton (2014), An overview of dissolved Fe and Mn Distributions during the 2010–2011 U.S. GEOTRACES north Atlantic Cruises: GEOTRACES GA03, *Deep Sea Research Part II: Topical Studies in Oceanography*. doi:10.1016/j.dsr2.2014.07.005

Hauck, J., C. Völker, T. Wang, M. Hoppema, M. Losch, and D. A. Wolf-Gladrow (2013), Seasonally different carbon flux changes in the Southern Ocean in response to the southern annular mode, *Global Biogeochemical Cycles*, *27*(4), 1236-1245. doi:10.1002/2013gb004600

Hayes, C. T., et al. (2015), Intensity of Th and Pa scavenging partitioned by particle chemistry in the North Atlantic Ocean, *Marine Chemistry*, *170*, 49-60. doi:10.1016/j.marchem.2015.01.006

Homoky, W. B., S. G. John, T. M. Conway, and R. A. Mills (2013), Distinct iron isotopic signatures and supply from marine sediment dissolution, *Nature communications*, *4*, 2143. doi:10.1038/ncomms3143

Honeyman, B. D., and P. H. Santschi (1989), A Brownian-pumping model for oceanic trace metal scavenging: Evidence from Th isotopes, *Journal of Marine Research*, *47*(4), 951-992. doi:10.1357/002224089785076091

Honeyman, B. D., L. S. Balistrieri, and J. W. Murray (1988), Oceanic trace metal scavenging: the importance of particle concentration, *Deep Sea Research Part A. Oceanographic Research Papers*, *35*(2), 227-246. doi:10.1016/0198-0149(88)90038-6

Hutchins, D. A., and K. W. Bruland (1994), Grazer-Mediated Regeneration and Assimilation of Fe, Zn and Mn from Planktonic Prey, *Marine Ecology Progress Series*, *110*(2-3), 259-269. doi:Doi 10.3354/Meps110259

Jickells, T. D., et al. (2005), Global iron connections between desert dust, ocean biogeochemistry, and climate, *Science*, *308*(5718), 67-71. doi:10.1126/Science.1105959

887 Johnson, K. S., R. M. Gordon, and K. H. Coale (1997), What controls dissolved iron
888 concentrations in the world ocean?, *Marine Chemistry*, *57*(3-4), 137-161. doi:10.1016/s0304-
889 4203(97)00043-1

890 Klunder, M. B., P. Laan, R. Middag, H. J. W. De Baar, and J. C. van Ooijen (2011), Dissolved iron
891 in the Southern Ocean (Atlantic sector), *Deep Sea Research Part II: Topical Studies in*
892 *Oceanography*, *58*(25-26), 2678-2694. doi:10.1016/j.dsr2.2010.10.042

893 Kohfeld, K. E., and A. Ridgwell (2009), Glacial-Interglacial Variability in Atmospheric CO2,
894 *Surface Ocean--Lower Atmosphere Processes*, *187*, 251-286. doi:10.1029/2008gm000845

895 Lam, P. J., and O. Marchal (2015), Insights into particle cycling from thorium and particle data,
896 *Annual review of marine science*, *7*, 159-184. doi:10.1146/annurev-marine-010814-015623

897 Lefèvre, N., and A. J. Watson (1999), Modeling the geochemical cycle of iron in the oceans and
898 its impact on atmospheric CO2concentrations, *Global Biogeochemical Cycles*, *13*(3), 727-736.
899 doi:10.1029/1999gb900034

900 Liu, X., and F. J. Millero (1999), The solubility of iron hydroxide in sodium chloride solutions,
901 *Geochimica et Cosmochimica Acta*, *63*(19-20), 3487-3497. doi:10.1016/s0016-
902 7037(99)00270-7

903 Marchal, O., and P. J. Lam (2012), What can paired measurements of Th isotope activity and
904 particle concentration tell us about particle cycling in the ocean?, *Geochimica et Cosmochimica*
905 *Acta*, *90*, 126-148. doi:10.1016/j.gca.2012.05.009

906 Marchetti, A., M. S. Parker, L. P. Moccia, E. O. Lin, A. L. Arrieta, F. Ribalet, M. E. Murphy, M. T.
907 Maldonado, and E. V. Armbrust (2009), Ferritin is used for iron storage in bloom-forming
908 marine pennate diatoms, *Nature*, *457*(7228), 467-470. doi:10.1038/nature07539

909 Matsumoto, K., K. Tokos, A. Huston, and H. Joy-Warren (2013), MESMO 2: a mechanistic
910 marine silica cycle and coupling to a simple terrestrial scheme, *Geoscientific Model*
911 *Development*, *6*(2), 477-494. doi:10.5194/gmd-6-477-2013

912 Mawji, E., R. Schlitzer, E. Masferrer-Dodas, and GEOTRACES-group (2015), The GEOTRACES
913 Intermediate Data Product 2014, *Marine Chemistry*. doi:10.1016/j.marchem.2015.04.005

914 Misumi, K., K. Lindsay, J. K. Moore, S. C. Doney, D. Tsumune, and Y. Yoshida (2013), Humic
915 substances may control dissolved iron distributions in the global ocean: Implications from
916 numerical simulations, *Global Biogeochemical Cycles*, *27*(2), 450-462. doi:10.1002/gbc.20039

917 Moore, C. M., et al. (2013), Processes and patterns of oceanic nutrient limitation, *Nature*
918 *Geoscience*. doi:10.1038/ngeo1765

919 Moore, J. K., and O. Braucher (2008), Sedimentary and mineral dust sources of dissolved iron
920 to the world ocean, *Biogeosciences*, *5*(3), 631-656. doi:10.5194/bg-5-631-2008

921 Moore, J. K., S. C. Doney, D. M. Glover, and I. Y. Fung (2002), Iron cycling and nutrient-
922 limitation patterns in surface waters of the World Ocean, *Deep-Sea Res Pt Ii*, *49*(1-3), 463-507.
923 doi:10.1016/S0967-0645(01)00109-6

924 Moore, J. K., K. Lindsay, S. C. Doney, M. C. Long, and K. Misumi (2013), Marine Ecosystem
925 Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]:

926 Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 Scenarios, *Journal of*
927 *Climate*, *26*(23), 9291-9312. doi:10.1175/jcli-d-12-00566.1

928 Noble, A. E., et al. (2012), Basin-scale inputs of cobalt, iron, and manganese from the
929 Benguela-Angola front to the South Atlantic Ocean, *Limnology and Oceanography*, *57*(4), 989-
930 1010. doi:10.4319/lo.2012.57.4.0989

931 Parekh, P., M. J. Follows, and E. Boyle (2004), Modeling the global ocean iron cycle, *Global*
932 *Biogeochemical Cycles*, *18*(1). doi:10.1029/2003gb002061

933 Raiswell, R., L. G. Benning, M. Tranter, and S. Tulaczyk (2008), Bioavailable iron in the
934 Southern Ocean: the significance of the iceberg conveyor belt, *Geochemical transactions*, *9*, 7.
935 doi:10.1186/1467-4866-9-7

936 Raven, J. A. (1988), The iron and molybdenum use efficiencies of plant growth with different
937 energy, carbon and nitrogen sources, *New Phytologist*, *109*(3), 279-287. doi:10.1111/j.1469-
938 8137.1988.tb04196.x

939 Raven, J. A., M. C. W. Evans, and R. E. Korb (1999), The role of trace metals in photosynthetic
940 electron transport in O2-evolving organisms, *Photosynthesis research*, *60*(2/3), 111-150.
941 doi:10.1023/a:1006282714942

942 Resing, J. A., P. N. Sedwick, C. R. German, W. Jenkins, J. W. Moffett, B. Sohst, and A. Tagliabue
943 (2015), Basin-scale transport of hydrothermal dissolved metals across the South Pacific
944 Ocean, *Nature*. doi:10.1038/nature14577

945 Ridgwell, A. J. (2001), Glacial-interglacial perturbations in the global carbon cycle, University
946 of East Anglia.

947 Rijkenberg, M. J., R. Middag, P. Laan, L. J. Gerringa, H. M. van Aken, V. Schoemann, J. T. de Jong,
948 and H. J. de Baar (2014), The distribution of dissolved iron in the west atlantic ocean, *Plos One*,
949 *9*(6), e101323. doi:10.1371/journal.pone.0101323

950 Saito, M. A., A. E. Noble, A. Tagliabue, T. J. Goepfert, C. H. Lamborg, and W. J. Jenkins (2013),
951 Slow-spreading submarine ridges in the South Atlantic as a significant oceanic iron source,
952 *Nature Geoscience*, *6*(9), 775-779. doi:10.1038/Ngeo1893

953 Sarthou, G., D. Vincent, U. Christaki, I. Obernosterer, K. R. Timmermans, and C. P. D. Brussaard
954 (2008), The fate of biogenic iron during a phytoplankton bloom induced by natural
955 fertilisation: Impact of copepod grazing, *Deep Sea Research Part II: Topical Studies in*
956 *Oceanography*, *55*(5-7), 734-751. doi:10.1016/j.dsr2.2007.12.033

957 Savoye, N., C. Benitez-Nelson, A. B. Burd, J. K. Cochran, M. Charette, K. O. Buesseler, G. A.
958 Jackson, M. Roy-Barman, S. Schmidt, and M. Elskens (2006), 234Th sorption and export
959 models in the water column: A review, *Marine Chemistry*, *100*(3-4), 234-249.
960 doi:10.1016/j.marchem.2005.10.014

961 Stock, C. A., J. P. Dunne, and J. G. John (2014), Global-scale carbon and energy flows through
962 the marine planktonic food web: An analysis with a coupled physical–biological model,
963 *Progress in Oceanography*, *120*, 1-28. doi:10.1016/j.pocean.2013.07.001

964 Strzepek, R. F., K. A. Hunter, R. D. Frew, P. J. Harrison, and P. W. Boyd (2012), Iron-light
965 interactions differ in Southern Ocean phytoplankton, *Limnology and Oceanography*, *57*(4),
966 1182-1200. doi:10.4319/lo.2012.57.4.1182

967 Strzepek, R. F., M. T. Maldonado, J. L. Higgins, J. Hall, K. Safi, S. W. Wilhelm, and P. W. Boyd
968 (2005), Spinning the "Ferrous Wheel": The importance of the microbial community in an iron
969 budget during the FeCycle experiment, *Global Biogeochemical Cycles*, *19*(4), GB4S26.
970 doi:10.1029/2005gb002490

971 Sunda, W. G., and S. A. Huntsman (1997), Interrelated influence of iron, light and cell size on
972 marine phytoplankton growth, *Nature*, *390*(6658), 389-392. doi:10.1038/37093

973 Sunda, W. G., and S. A. Huntsman (1998), Processes regulating cellular metal accumulation
974 and physiological effects: Phytoplankton as model systems, *Science of The Total Environment*,
975 *219*(2-3), 165-181. doi:10.1016/s0048-9697(98)00226-5

976 Tagliabue, A. (2014), More to hydrothermal iron input than meets the eye, *Proceedings of the*
977 *National Academy of Sciences of the United States of America*, *111*(47), 16641-16642.
978 doi:10.1073/pnas.1419829111

979 Tagliabue, A., and C. Völker (2011), Towards accounting for dissolved iron speciation in global
980 ocean models, *Biogeosciences*, *8*(10), 3025-3039. doi:10.5194/bg-8-3025-2011

981 Tagliabue, A., L. Bopp, and O. Aumont (2008), Ocean biogeochemistry exhibits contrasting
982 responses to a large scale reduction in dust deposition, *Biogeosciences*, *5*(1), 11-24.

983 Tagliabue, A., O. Aumont, and L. Bopp (2014a), The impact of different external sources of iron
984 on the global carbon cycle, *Geophysical Research Letters*, *41*(3), 920-926.
985 doi:10.1002/2013gl059059

986 Tagliabue, A., R. G. Williams, N. Rogan, E. P. Achterberg, and P. W. Boyd (2014b), A ventilation-
987 based framework to explain the regeneration-scavenging balance of iron in the ocean,
988 *Geophysical Research Letters*, *41*(20), 7227-7236. doi:10.1002/2014gl061066

989 Tagliabue, A., J.-B. Sallée, A. R. Bowie, M. Lévy, S. Swart, and P. W. Boyd (2014c), Surface-water
990 iron supplies in the Southern Ocean sustained by deep winter mixing, *Nature Geoscience*, *7*(4),
991 314-320. doi:10.1038/ngeo2101

992 Tagliabue, A., T. Mtshali, O. Aumont, A. R. Bowie, M. B. Klunder, A. N. Roychoudhury, and S.
993 Swart (2012), A global compilation of dissolved iron measurements: focus on distributions
994 and processes in the Southern Ocean, *Biogeosciences*, *9*(6), 2333-2349. doi:10.5194/bg-9-
995 2333-2012

996 Tagliabue, A., L. Bopp, D. M. Roche, N. Bouttes, J. C. Dutay, R. Alkama, M. Kageyama, E. Michel,
997 and D. Paillard (2009), Quantifying the roles of ocean circulation and biogeochemistry in
998 governing ocean carbon-13 and atmospheric carbon dioxide at the last glacial maximum,
999 *Climate of the Past*, *5*(4), 695-706. doi:10.5194/cp-5-695-2009

1000 Tagliabue, A., et al. (2010), Hydrothermal contribution to the oceanic dissolved iron
1001 inventory, *Nature Geoscience*, *3*(4), 252-256. doi:10.1038/ngeo818

1002 Twining, B. S., and S. B. Baines (2013), The trace metal composition of marine phytoplankton,
1003 *Annual review of marine science*, *5*, 191-215. doi:10.1146/annurev-marine-121211-172322

1004 Twining, B. S., S. D. Nodder, A. L. King, D. A. Hutchins, G. R. LeCleir, J. M. DeBruyn, E. W. Maas, S.
1005 Vogt, S. W. Wilhelm, and P. W. Boyd (2014), Differential remineralization of major and trace
1006 elements in sinking diatoms, *Limnol. Oceanogr*, *59*(3), 689-704.
1007 doi:10.4319/lo.2014.59.3.0689

1008 Vichi, M., N. Pinardi, and S. Masina (2007), A generalized model of pelagic biogeochemistry for
1009 the global ocean ecosystem. Part I: Theory, *Journal of Marine Systems*, *64*(1-4), 89-109.
1010 doi:10.1016/j.jmarsys.2006.03.006

1011 Völker, C., and A. Tagliabue (2015), Modeling organic iron-binding ligands in a three-
1012 dimensional biogeochemical ocean model, *Marine Chemistry*, *173*, 67-77.
1013 doi:10.1016/j.marchem.2014.11.008

1014 Wagener, T., E. Pulido-Villena, and C. Guieu (2008), Dust iron dissolution in seawater: Results
1015 from a one-year time-series in the Mediterranean Sea, *Geophysical Research Letters*, *35*(16).
1016 doi:10.1029/2008gl034581

1017 Ward, B. A., S. Dutkiewicz, C. M. Moore, and M. J. Follows (2013), Iron, phosphorus, and
1018 nitrogen supply ratios define the biogeography of nitrogen fixation, *Limnology and*
1019 *Oceanography*, *58*(6), 2059-2075. doi:10.4319/lo.2013.58.6.2059

1020 Watson, A. J., D. C. E. Bakker, A. J. Ridgwell, P. W. Boyd, and C. S. Law (2000), Effect of iron
1021 supply on Southern Ocean CO2 uptake and implications for glacial atmospheric CO2, *Nature*,
1022 *407*(6805), 730-733. doi:10.1038/35037561

1023 Wells, M. L., N. M. Price, and K. W. Bruland (1995), Iron chemistry in seawater and its
1024 relationship to phytoplankton: a workshop report, *Marine Chemistry*, *48*(2), 157-182.
1025 doi:10.1016/0304-4203(94)00055-i

1026 Wu, J., E. Boyle, W. Sunda, and L. S. Wen (2001), Soluble and colloidal iron in the oligotrophic
1027 North Atlantic and North Pacific, *Science*, *293*(5531), 847-849. doi:10.1126/science.1059251

1028 Ye, Y., T. Wagener, C. Völker, C. Guieu, and D. A. Wolf-Gladrow (2011), Dust deposition: iron
1029 source or sink? A case study, *Biogeosciences*, *8*(8), 2107-2124. doi:10.5194/bg-8-2107-2011

1030 Yool, A., E. E. Popova, and T. R. Anderson (2011), Medusa-1.0: a new intermediate complexity
1031 plankton ecosystem model for the global domain, *Geoscientific Model Development*, *4*(2), 381-
1032 417. doi:10.5194/gmd-4-381-2011

1033 Yool, A., E. E. Popova, and T. R. Anderson (2013), MEDUSA-2.0: an intermediate complexity
1034 biogeochemical model of the marine carbon cycle for climate change and ocean acidification
1035 studies, *Geoscientific Model Development*, *6*(5), 1767-1811. doi:10.5194/gmd-6-1767-2013
1036
1037 **Figure Legends**
1038
1039 **Figure 1.** Histograms of the average DFe concentration (nM) simulated by the FeMIP models
1040 across four different depth bins for three regions. The Northern Hemisphere is 30°N-90°N,
1041 Tropics are 30°S-30°N and the Southern Hemisphere is 30°S-90°S.
1042
1043 **Figure 2.** Annual mean DFe concentrations (nM) averaged over the upper 50m from the
1044 FeMIP models. Data averaged over the period January to June and July to December is taken
1045 from the expanded *Tagliabue et al.* [2012] dataset and has been averaged over 5^{o} bins in
1046 latitude and longitude to improve visibility.

Figure 3. Annual maximum minus annual minimum DFe concentrations (nM) averaged over the upper 50m from the FeMIP models.

Figure 4. DFe concentrations (nM) from the GA-02 [*Rijkenberg et al.*, 2014] cruise and extracted from the FeMIP models.

Figure 5. DFe concentrations (nM) from the GA-03 cruise [*Hatta et al.*, 2014] and extracted from the FeMIP models.

Figure 6. DFe concentrations (nM) from the CoFeMUG cruise [*Noble et al.*, 2012] and extracted from the FeMIP models.

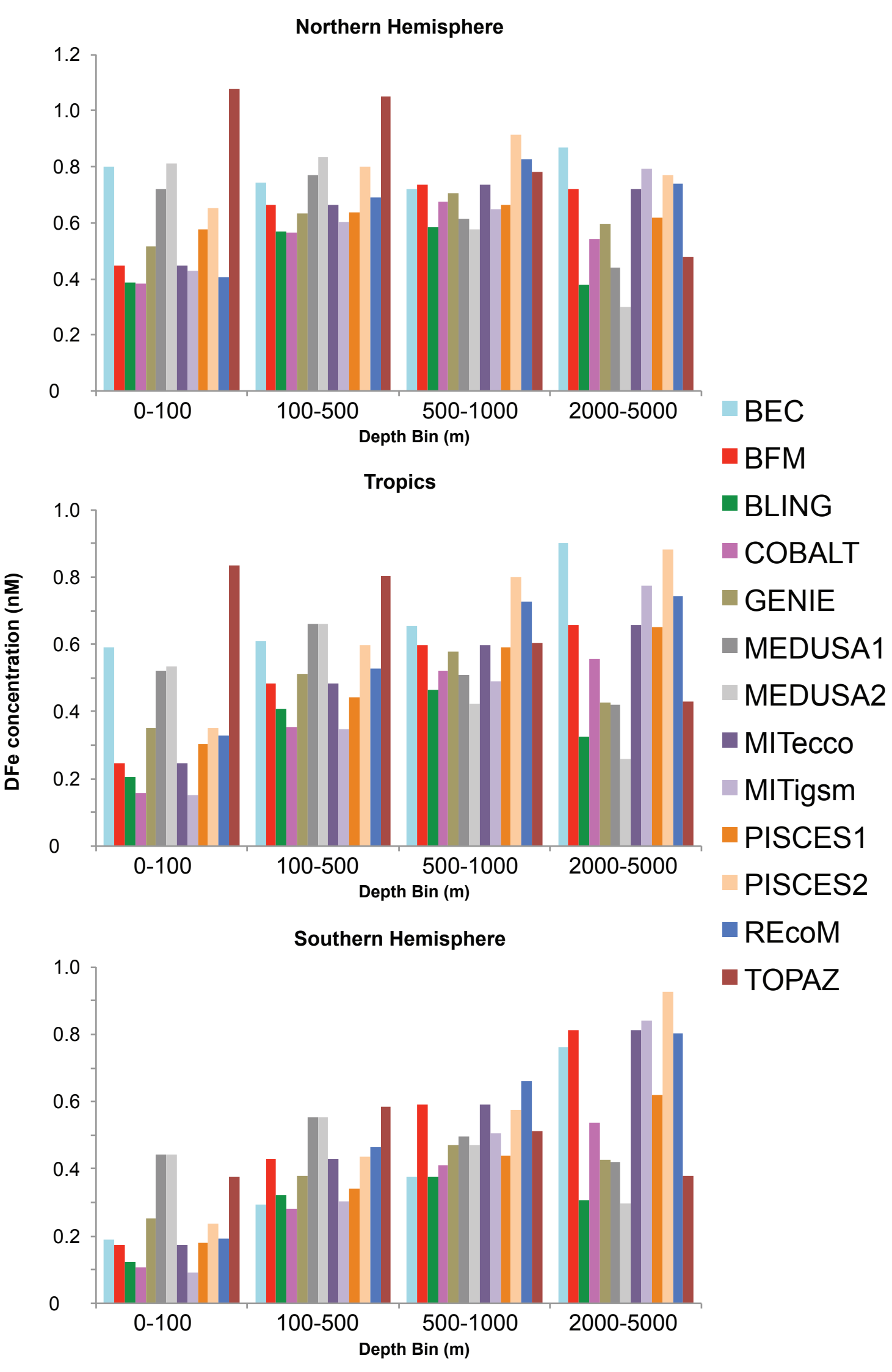
Figure 7. DFe concentrations (nM) from the GP-16 cruise [*Resing et al.*, 2015] and extracted from the FeMIP models.

Figure 8. DFe concentrations (nM) from the GIPY-4 and 5 cruises [*Chever et al.*, 2010; *Klunder et al.*, 2011] and extracted from the FeMIP models

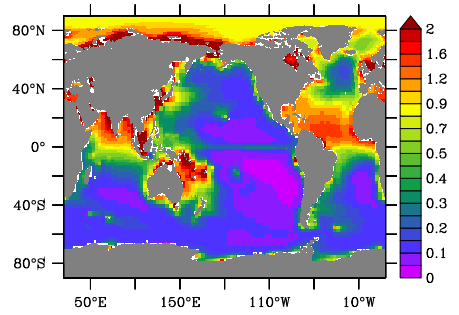
Figure 9. NO_3 concentrations (µM) from the GA-02 cruise [*Rijkenberg et al.*, 2014] and extracted from the FeMIP models (NO_3 data not provided for GENIE).

Figure 10. PO_4 concentrations (µM) from the GA-02 cruise [*Rijkenberg et al.*, 2014] and extracted from the FeMIP models (PO_4 not provided for MEDUSA-1, MEDUSA-2, RECOM and TOPAZ).

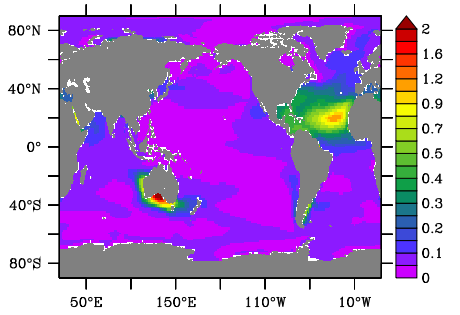
Figure 11. Fe* (Fe – NO_3*$r_{Fe/N}$, nM) from the GA-02 cruise cruise [*Rijkenberg et al.*, 2014] and extracted from the FeMIP models. For models that do not provide NO3, PO4 is used and converted to NO3 assuming a ratio of 16:1.


Northern Hemisphere
1.2
1.0
0.8
0.6
0.4
0.2
0
0-100
100-500
500-1000
2000-5000
Depth Bin (m)
Tropics
DFe concentration (nM)
1.0
0.8
0.6
0.4
0.2
0
0-100
100-500
500-1000
2000-5000
Depth Bin (m)
Southern Hemisphere
1.0
0.8
0.6
0.4
0.2
0
0-100
100-500
500-1000
2000-5000
Depth Bin (m)
BEC
BFM
BLING
COBALT
GENIE
MEDUSA1
MEDUSA2
MITecco
MITigsm
PISCES1
PISCES2
REcoM
TOPAZ

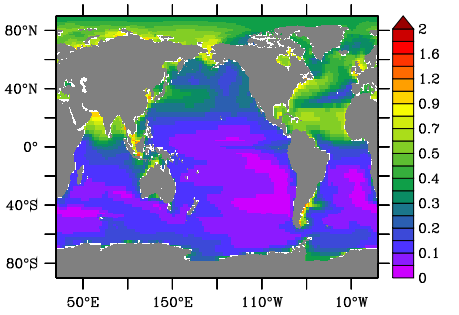
BEC



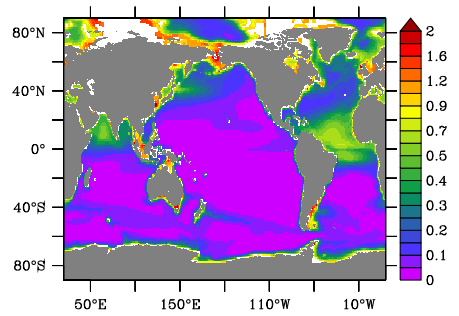
BFM



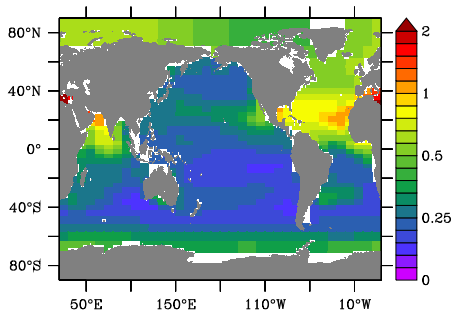
BLING



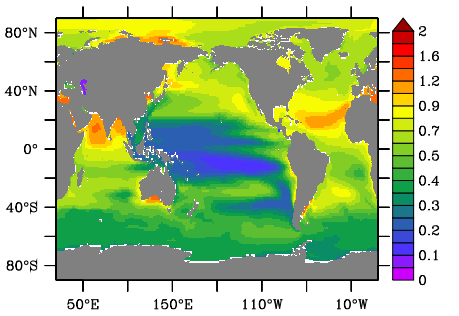
COBALT



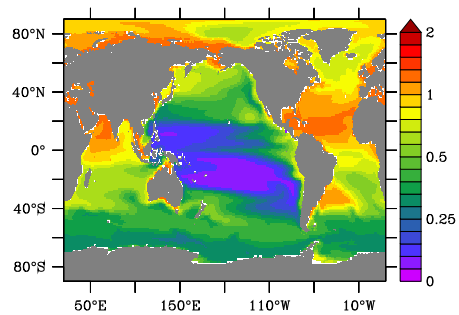
GENIE



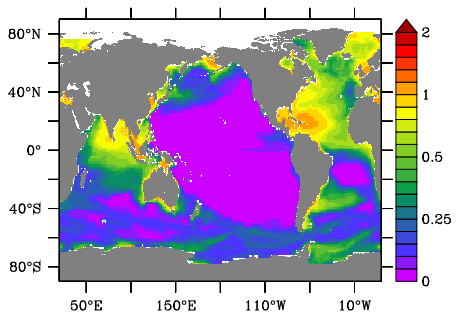
MEDUSA 1



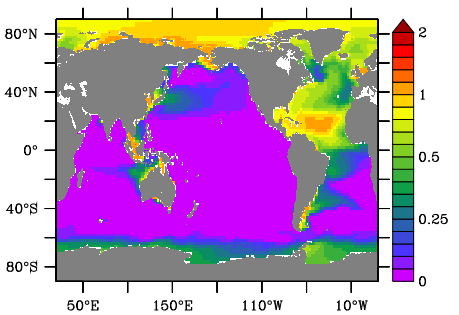
MEDUSA 2



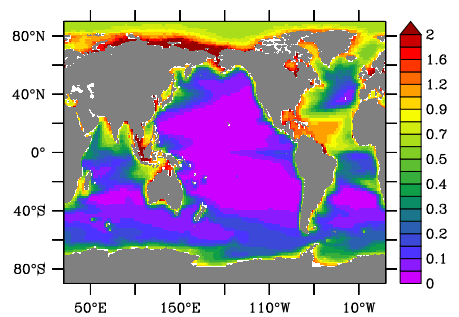
MIT ecco



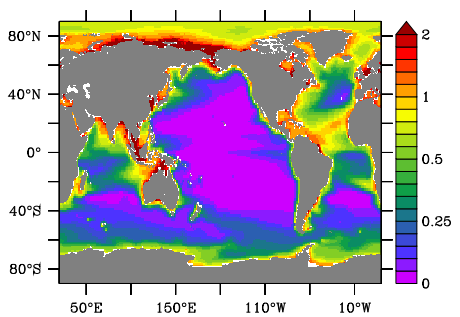
MIT igsm



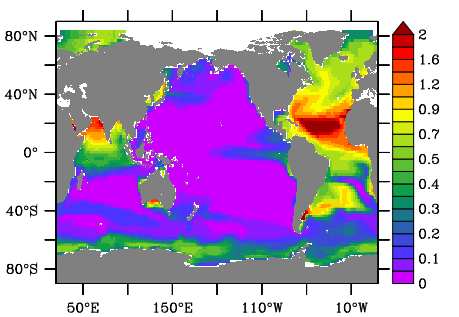
PISCES 1



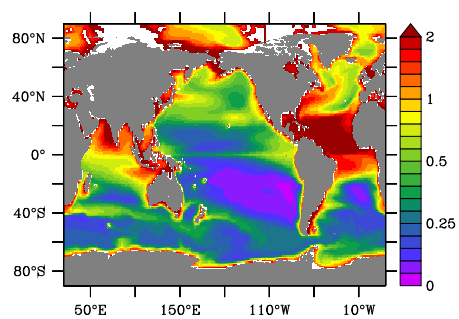
PISCES 2



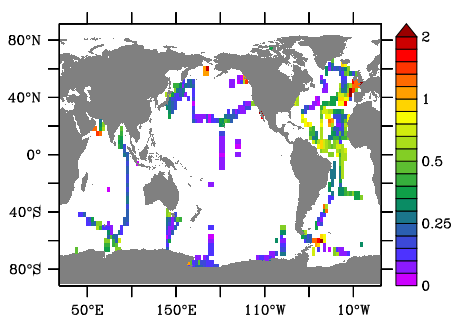
RECOM



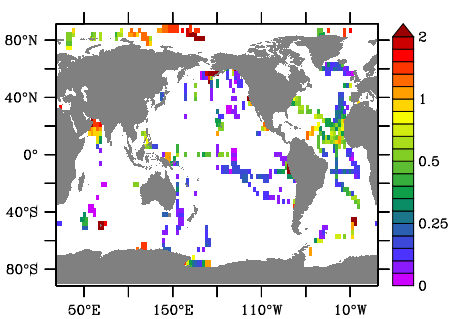
TOPAZ



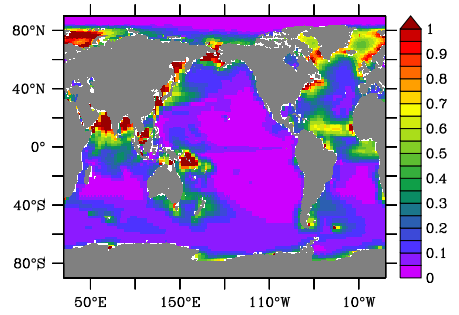
Data, Jan to Jun



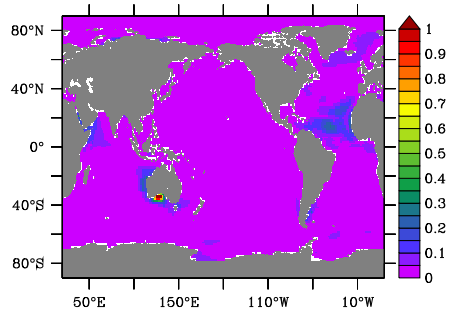
Data, Jul to Dec



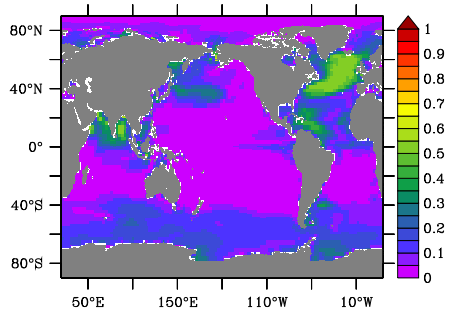
BEC



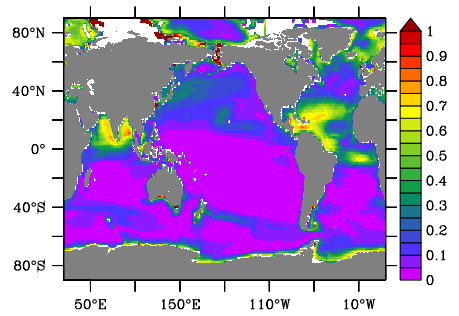
BFM



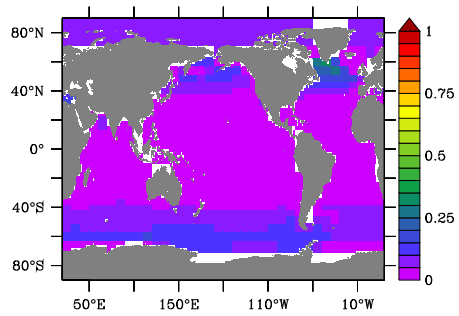
BLING



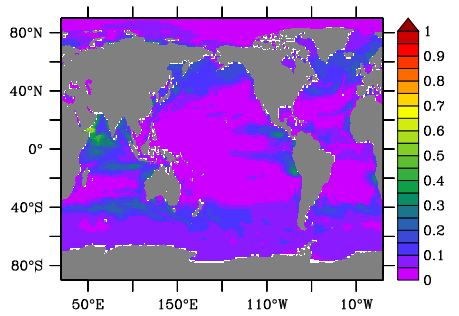
COBALT



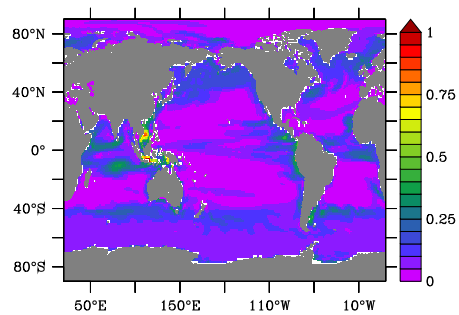
GENIE



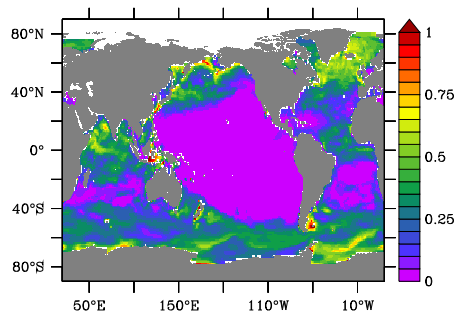
MEDUSA 1



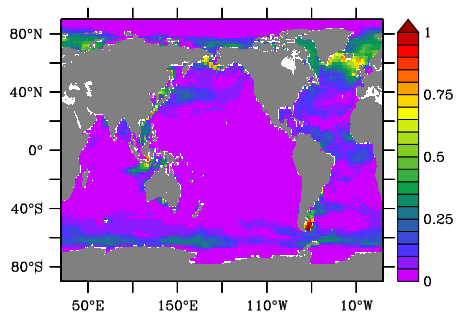
MEDUSA 2



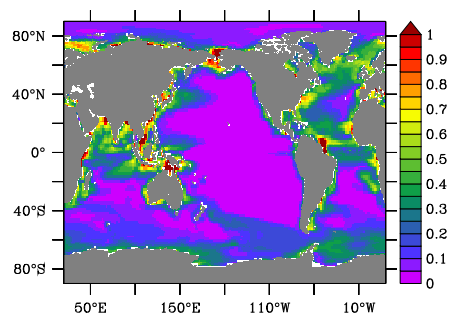
MIT ecco



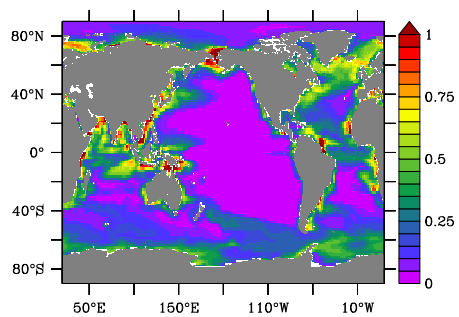
MIT igsm



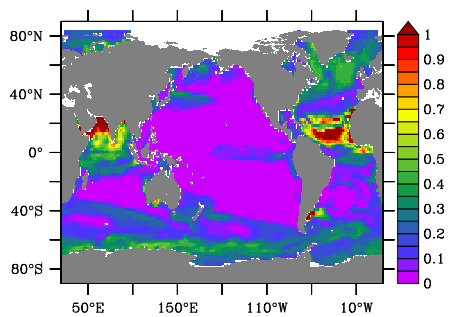
PISCES 1



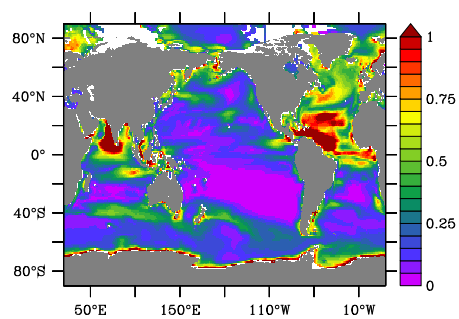
PISCES 2



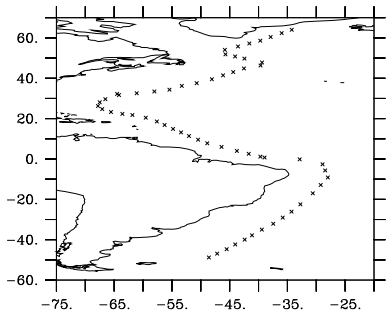
RECOM



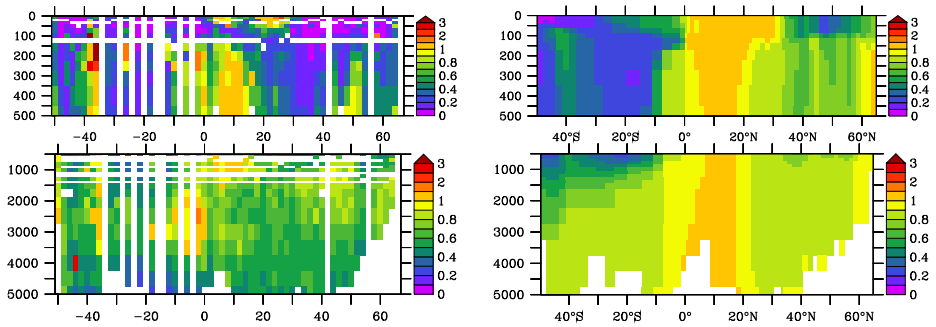
TOPAZ



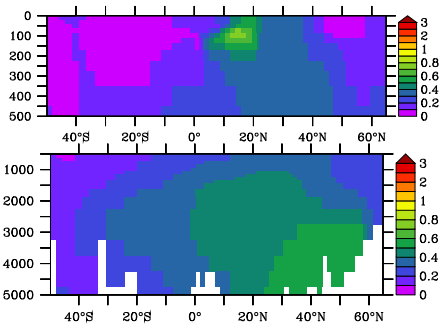
GA−02, Cruise



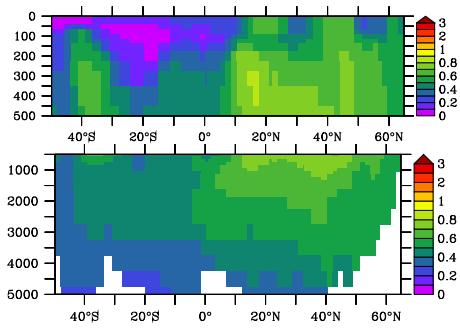
BEC, GA−02



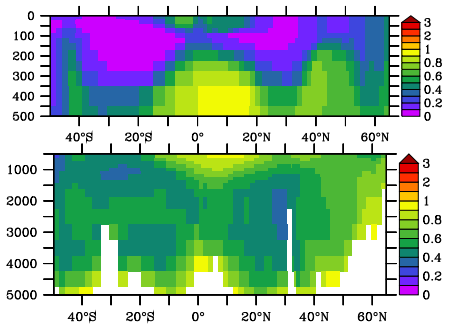
BFM, GA−02



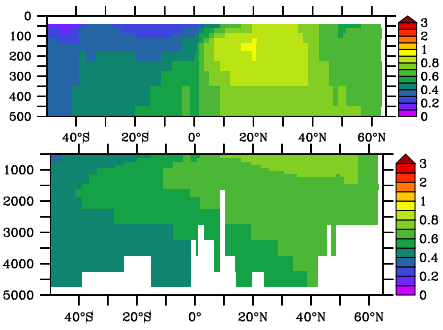
BLING, GA−02



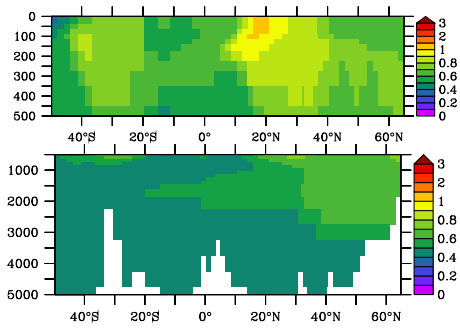
COBALT, GA−02



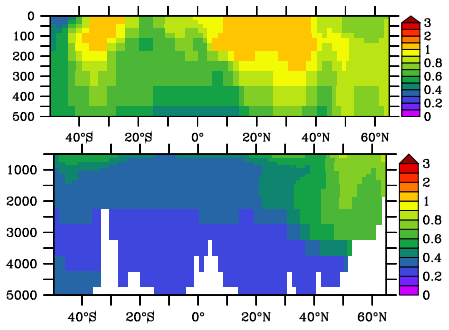
GENIE, GA−02



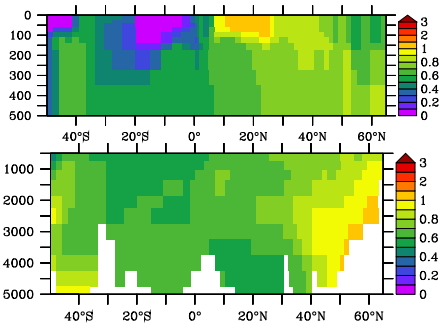
MEDUSA 1, GA−02



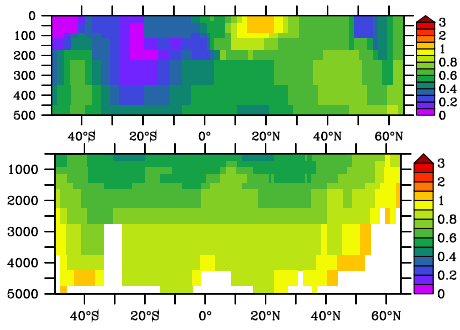
MEDUSA 2, GA−02



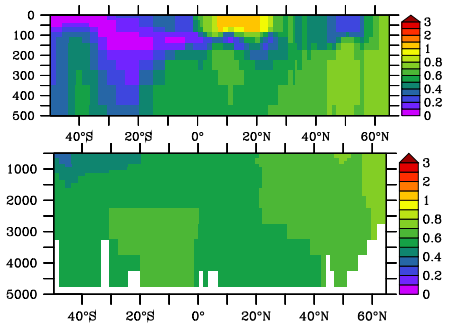
MITecco, GA−02



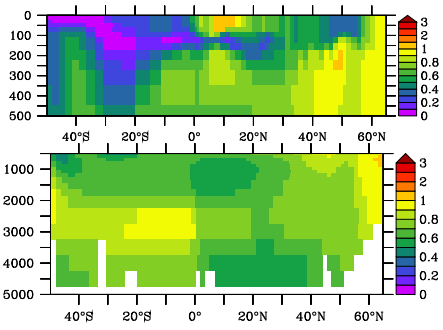
MITgsm, GA−02



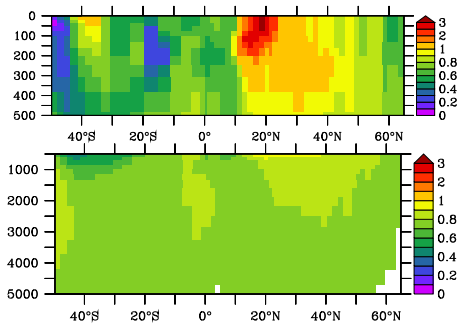
PISCES 1, GA−02



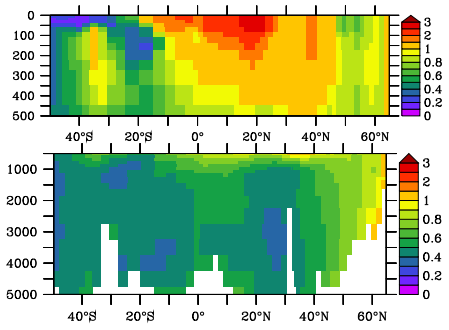
PISCES 2, GA−02



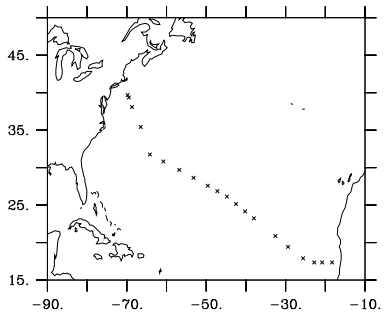
RECOM, GA−02

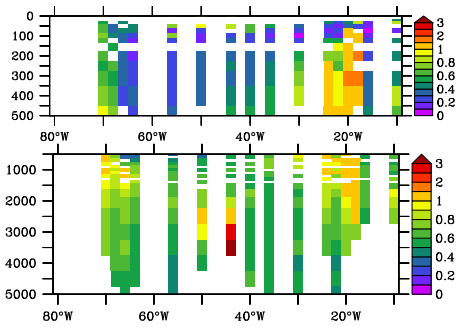


TOPAZ, GA−02

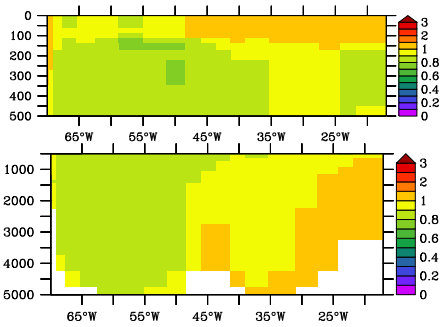


GA−03, Cruise

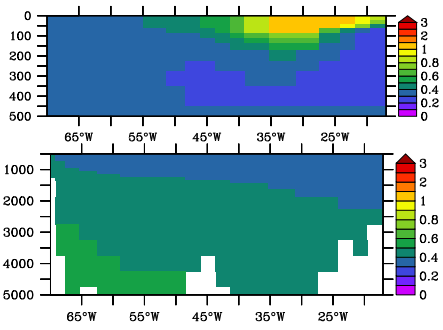




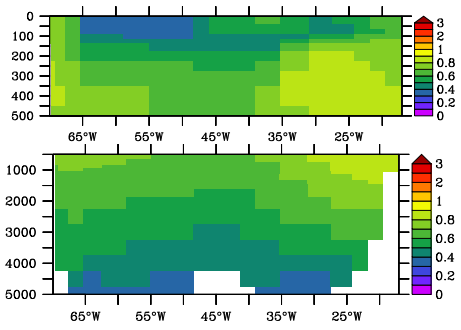
BEC, GA−03



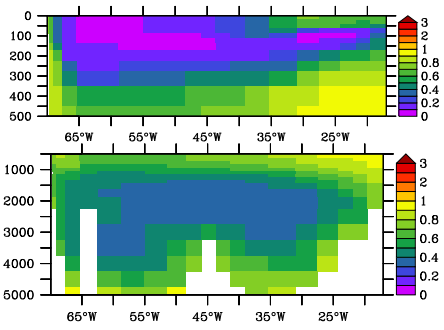
BFM, GA−03



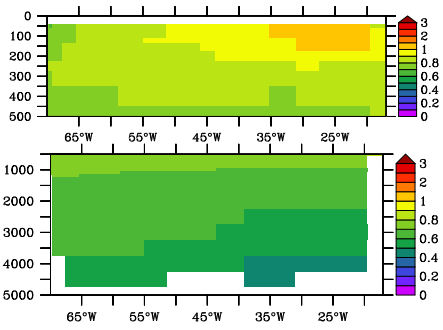
BLING, GA−03



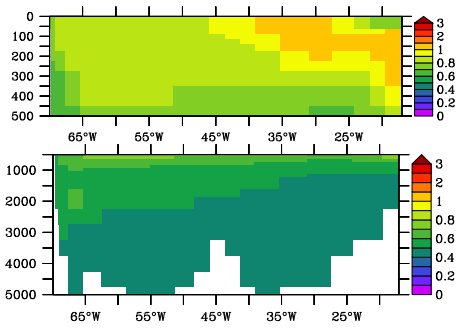
COBALT, GA−03



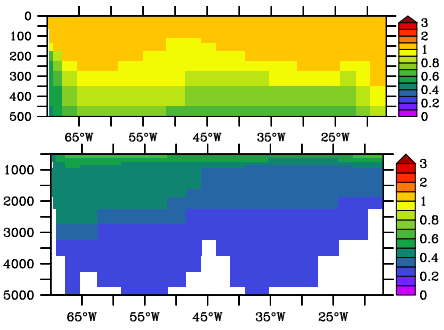
GENIE, GA−03



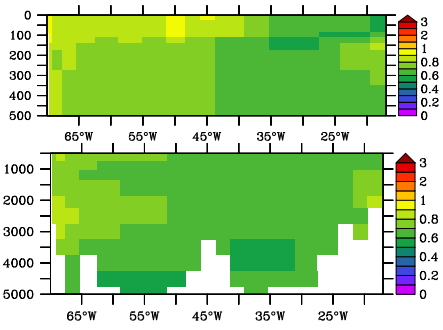
MEDUSA 1, GA−03



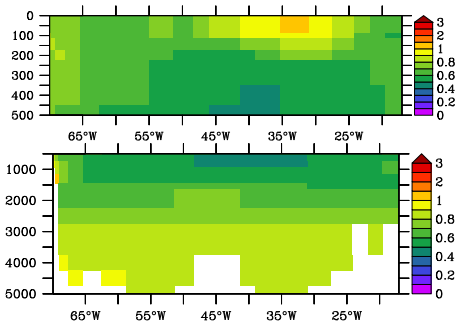
MEDUSA 2, GA−03



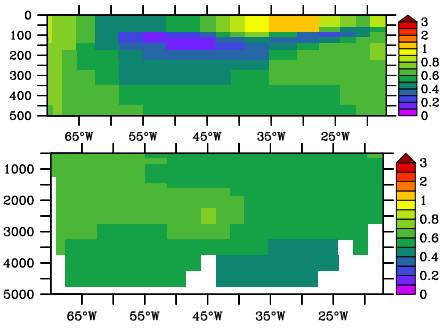
MITecco, GA−03



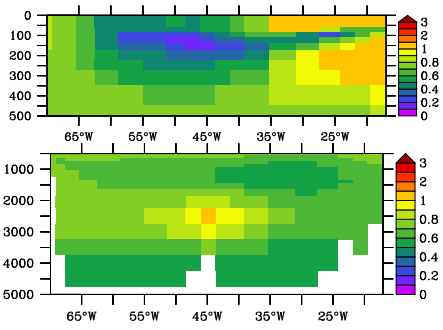
MITigsm, GA−03



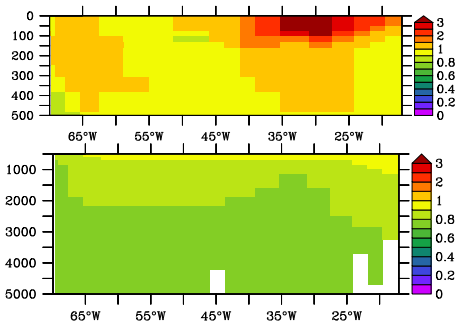
PISCES 1, GA−03



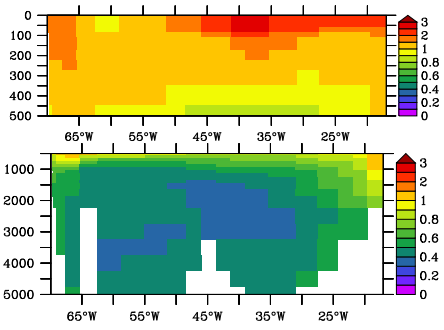
PISCES 2, GA−03



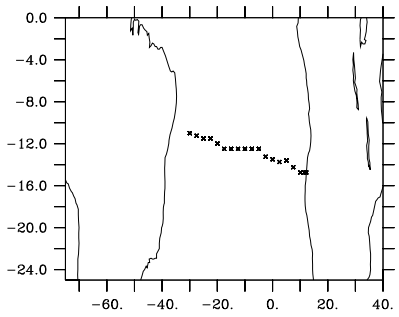
RECOM, GA−03

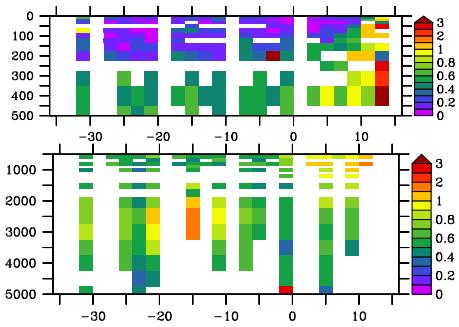


TOPAZ, GA−03

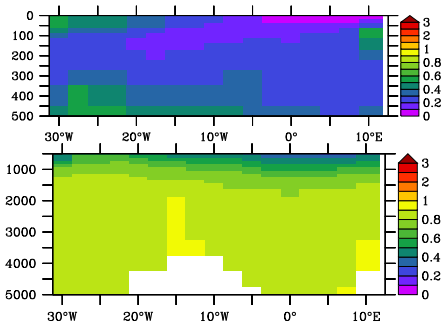


CoFeMUG, Cruise

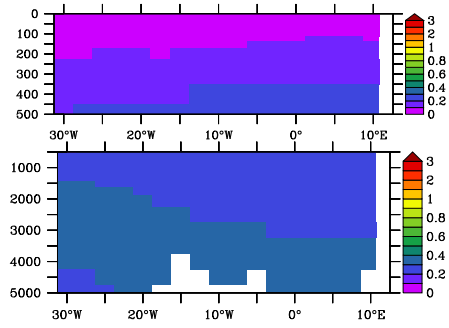




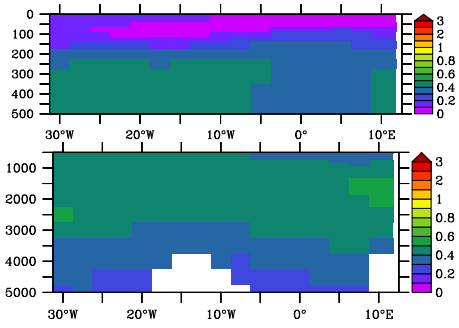
BEC, GA−03



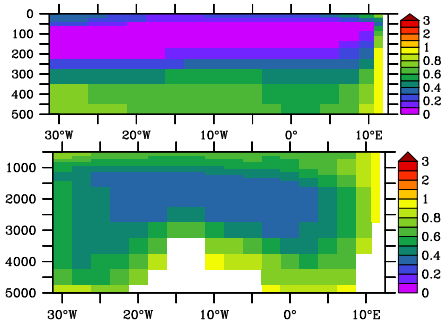
BFM, GA−03



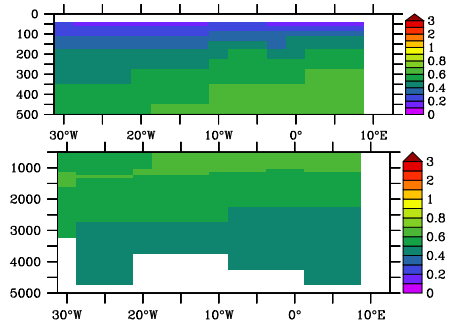
BLING, GA−03



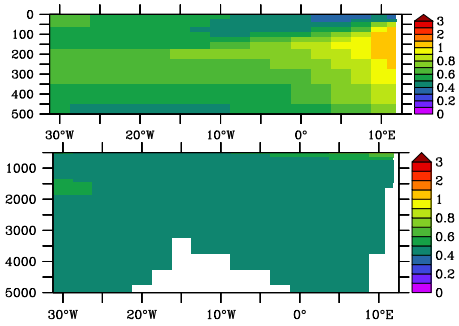
COBALT, GA−03



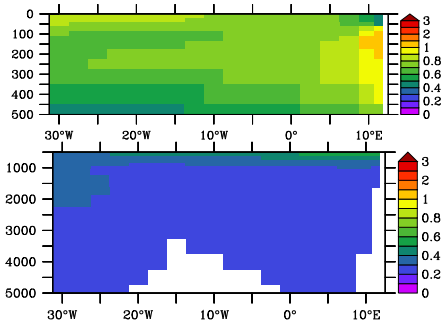
GENIE, GA−03



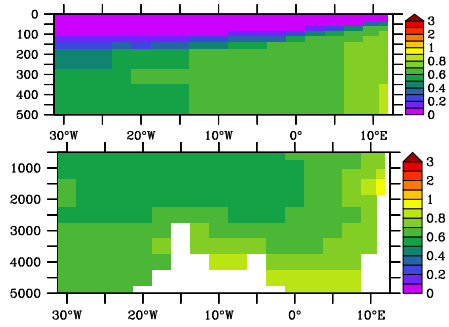
MEDUSA 1, GA−03



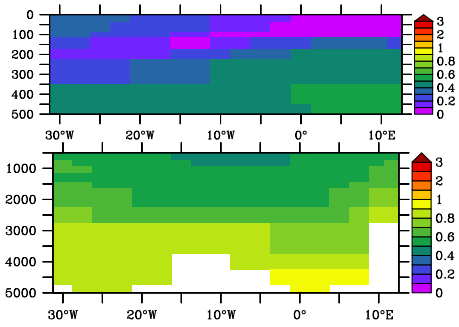
MEDUSA 2, GA−03



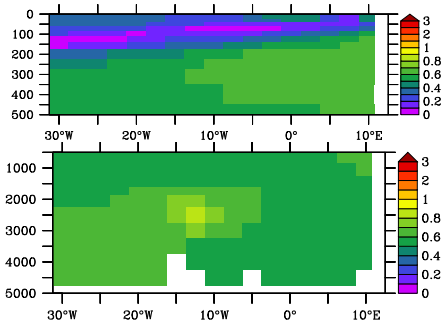
MITecco, GA−03



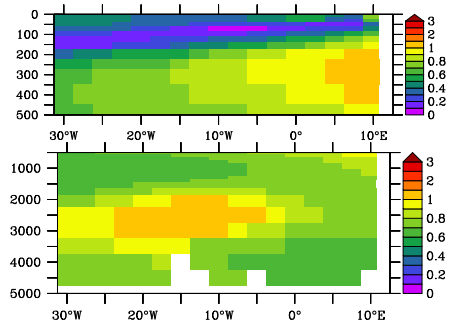
MITigsm, GA−03



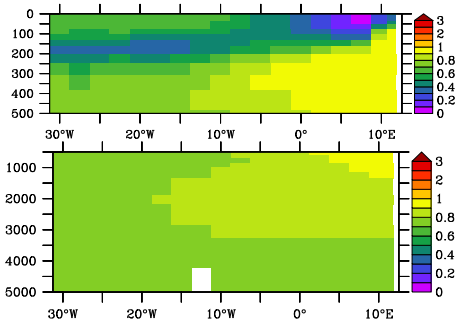
PISCES 1, GA−03



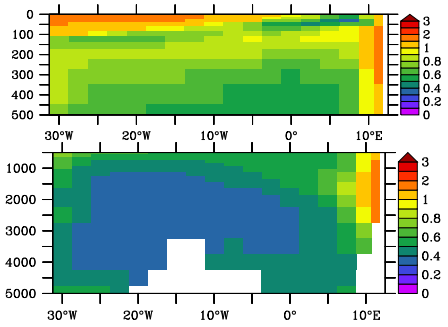
PISCES 2, GA−03



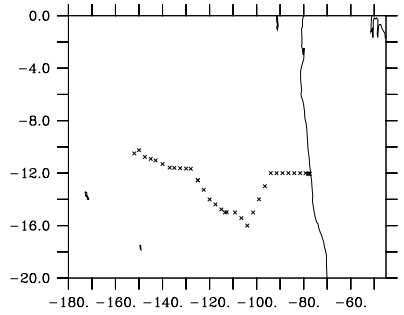
RECOM, GA−03



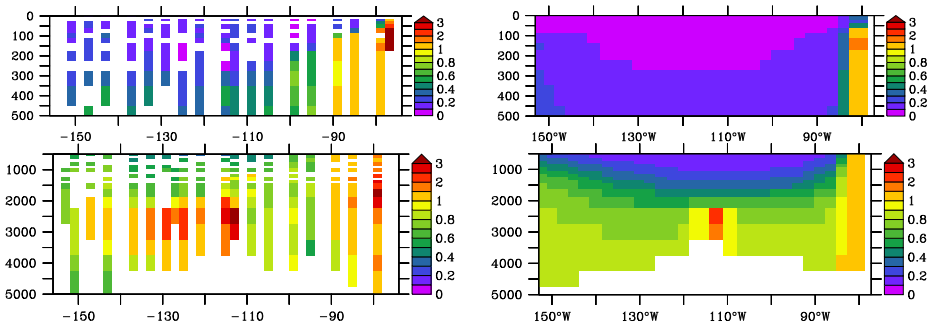
TOPAZ, GA−03



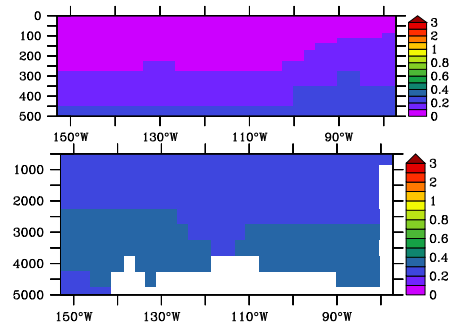
GP−16 Cruise



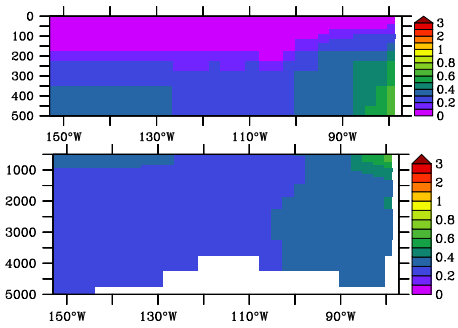
BEC, GP−16



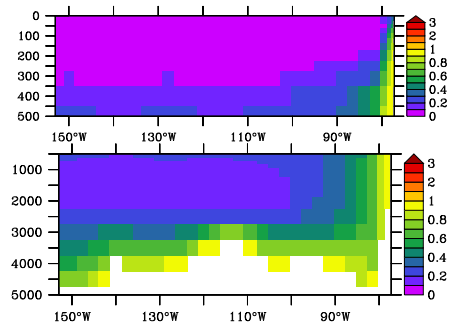
BFM, GP−16



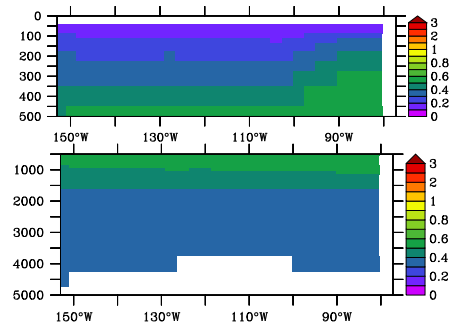
BLING, GP−16



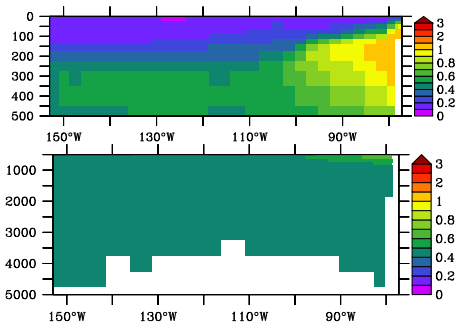
COBALT, GP−16



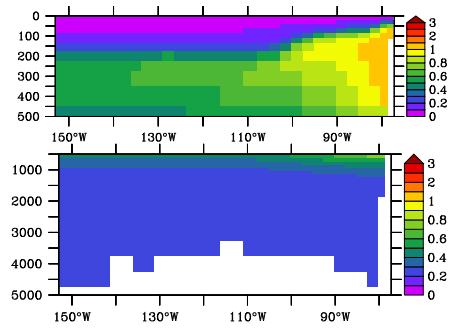
GENIE, GP−16



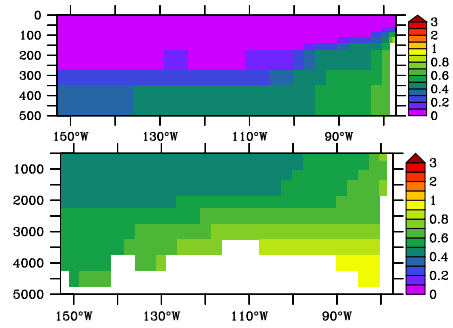
MEDUSA 1, GP−16



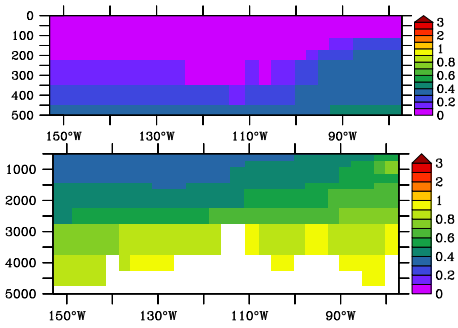
MEDUSA 2, GP−16



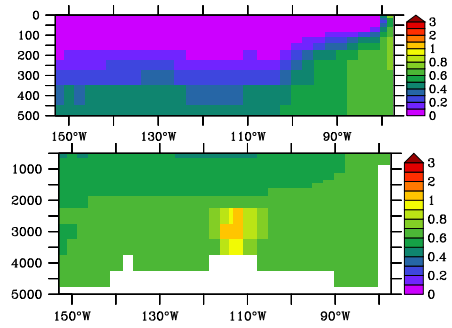
MITecco, GP−16



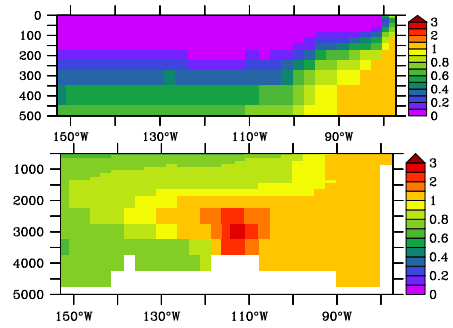
MITigsm, GP−16



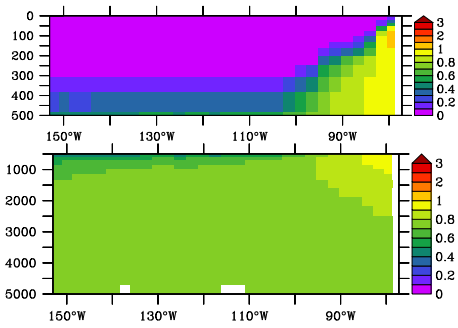
PISCES 1, GP−16



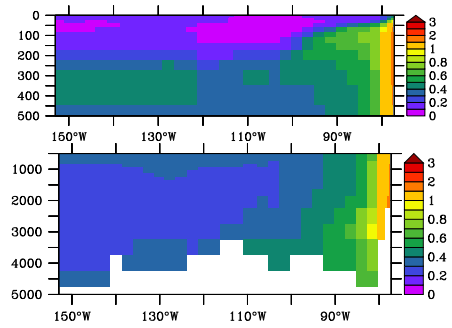
PISCES 2, GP−16



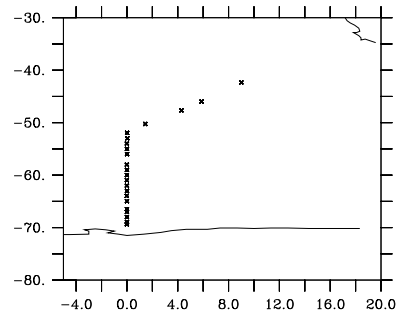
RECOM, GP−16

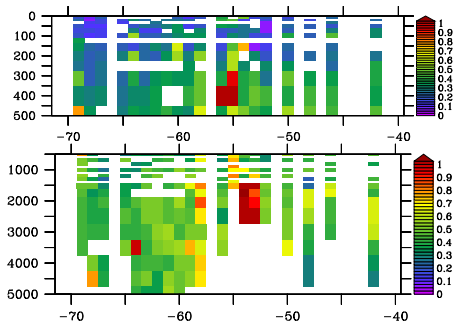


TOPAZ, GP−16

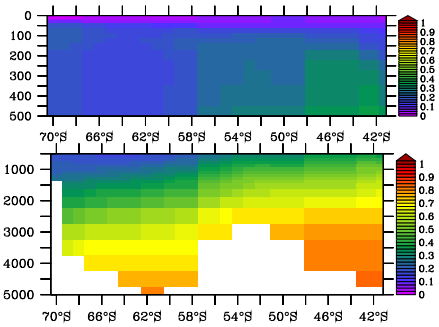


GIPY−5, Cruise

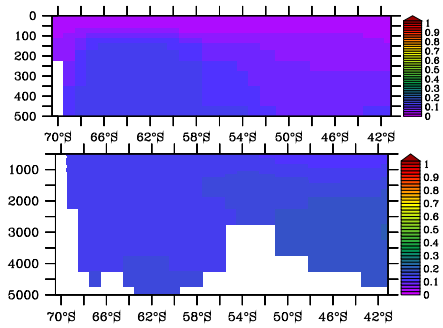




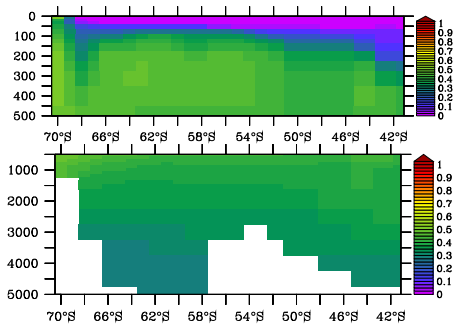
BEC, GIPY−5



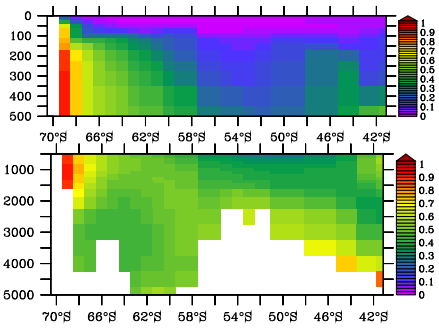
BFM, GIPY−5



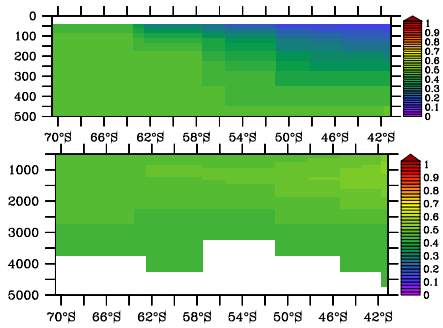
BLING, GIPY−5



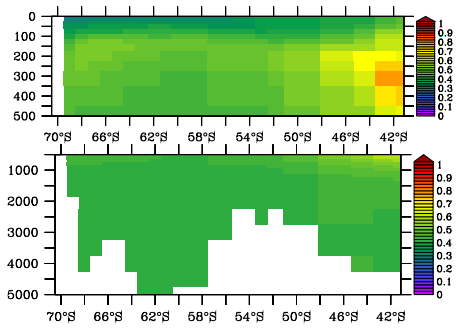
COBALT, GIPY−5



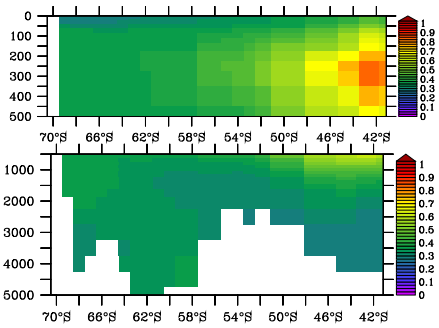
GENIE, GIPY−5



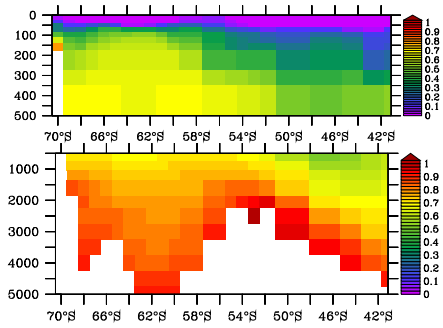
MEDUSA 1, GIPY−5



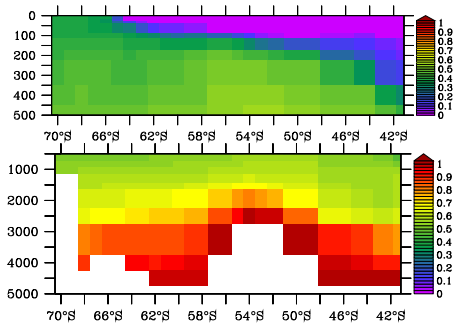
MEDUSA 2, GIPY−5



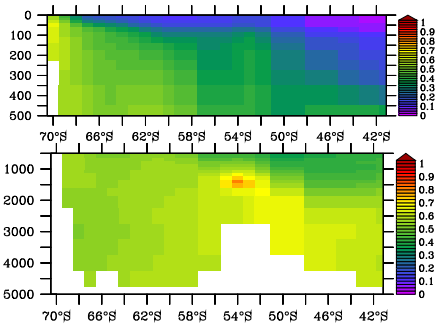
MITecco, GIPY−5



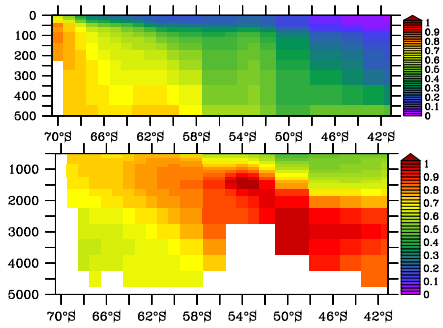
MITigsm, GIPY−5



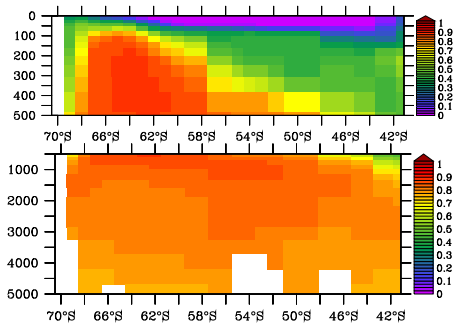
PISCES 1, GIPY−5



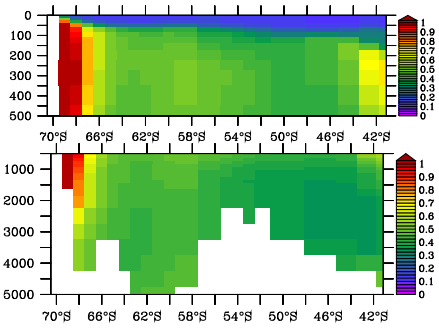
PISCES 2, GIPY−5



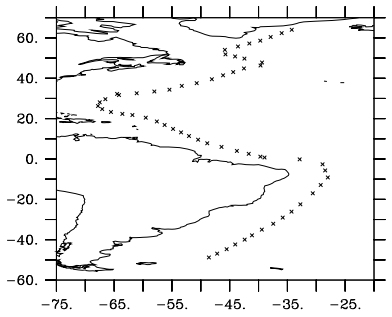
RECOM, GIPY−5



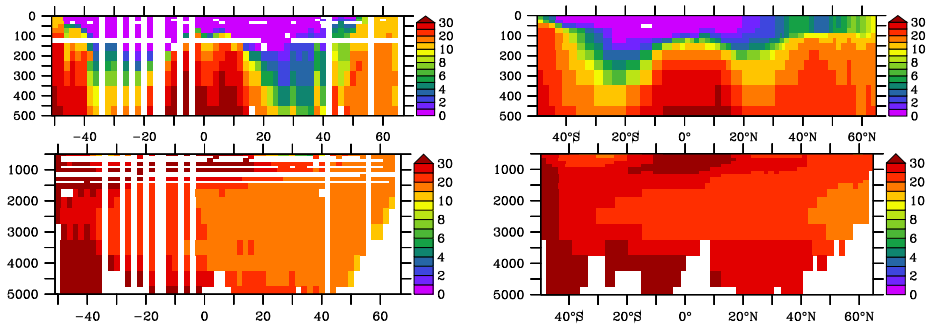
TOPAZ, GIPY−5



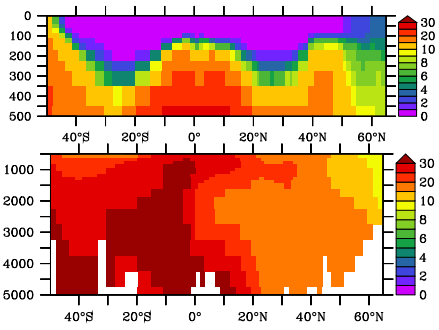
GA−02, Cruise



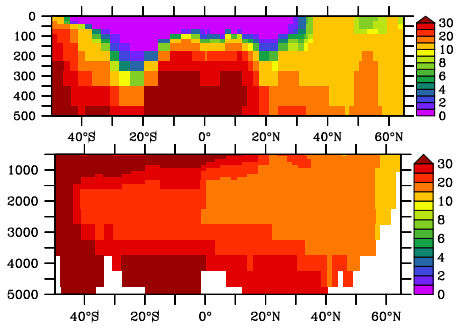
BEC, GA−02



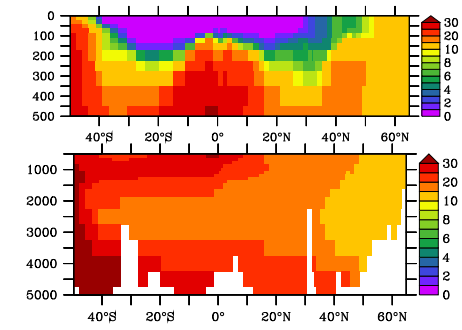
BFM, GA−02



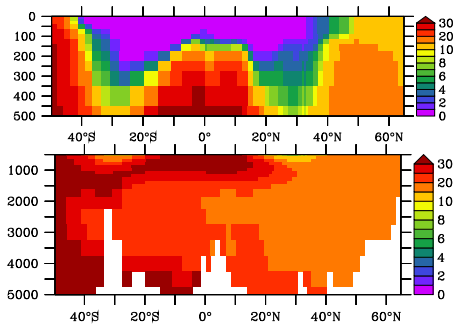
BLING, GA−02



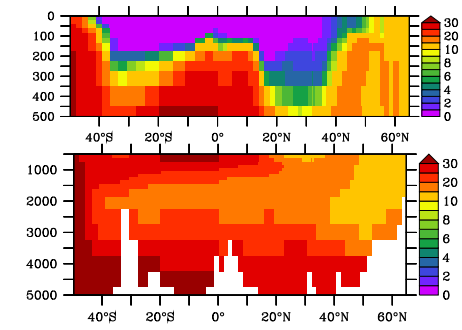
COBALT, GA−02



MEDUSA 1, GA−02

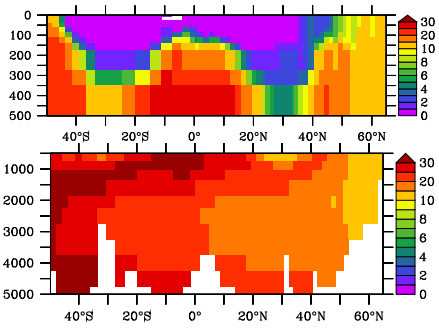


MEDUSA 2, GA−02

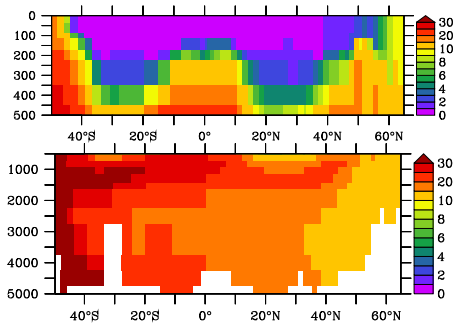


GENIE, GA−02

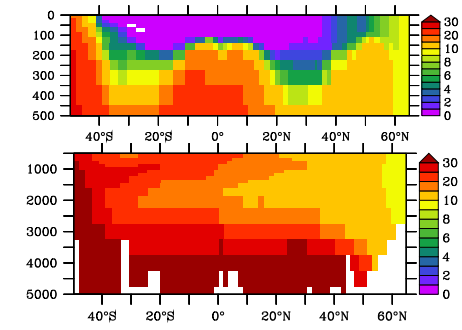
MITecco, GA−02



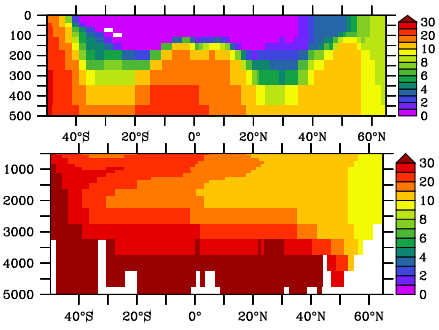
MITigsm, GA−02



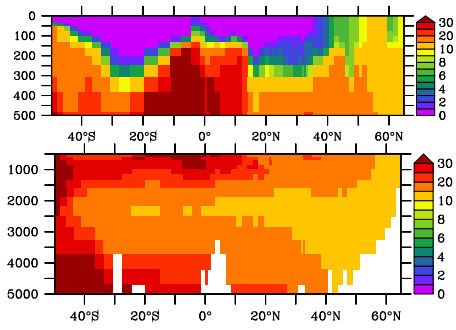
PISCES 1, GA−02



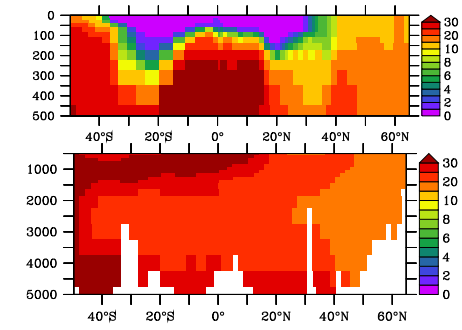
PISCES 2, GA−02



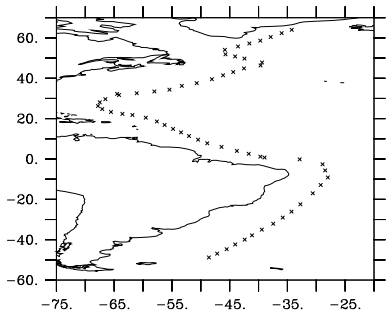
RECOM, GA−02



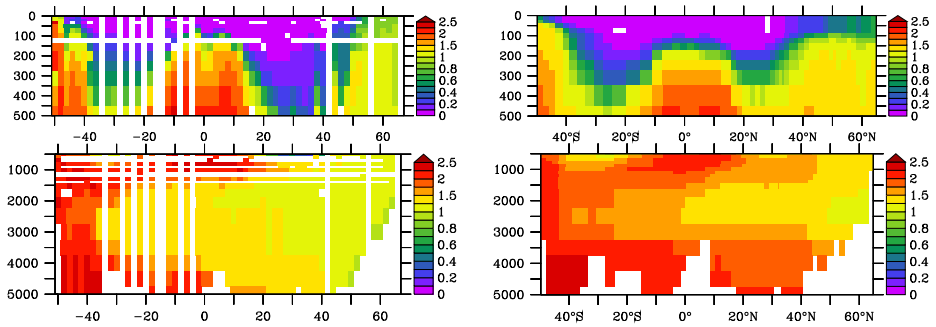
TOPAZ, GA−02



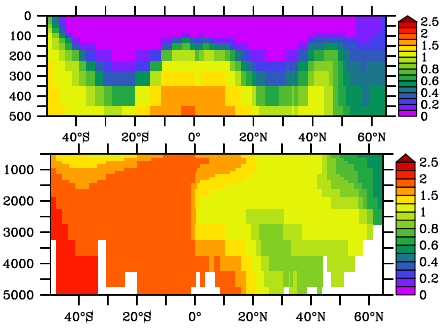
GA−02, Cruise



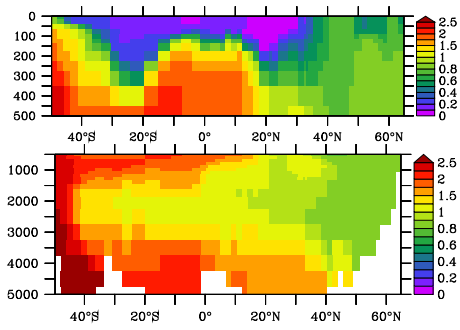
BEC, GA−02



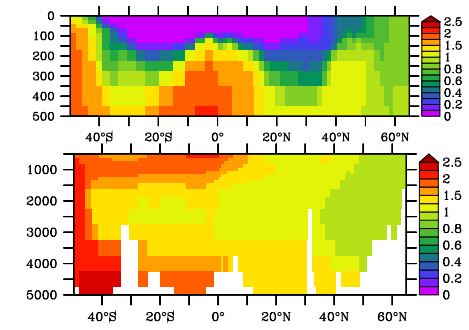
BFM, GA−02



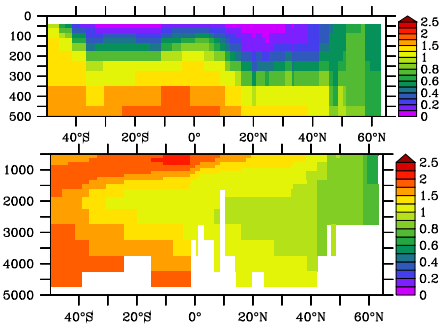
BLING, GA−02



COBALT, GA−02



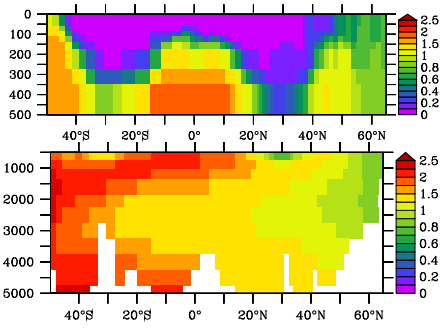
GENIE, GA−02



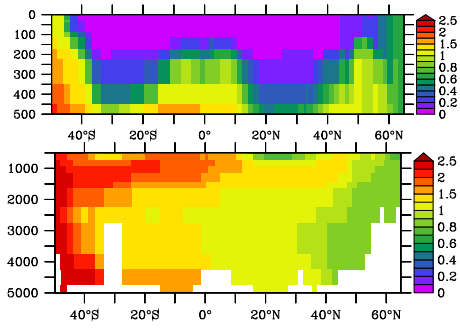
MEDUSA 1, GA−02

MEDUSA 2, GA−02

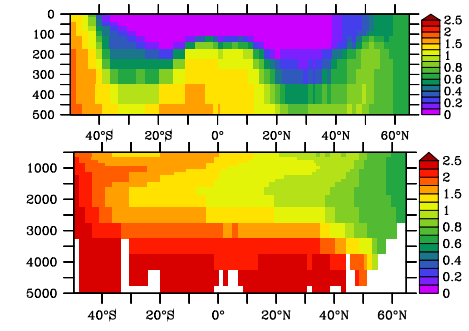
MITecco, GA−02



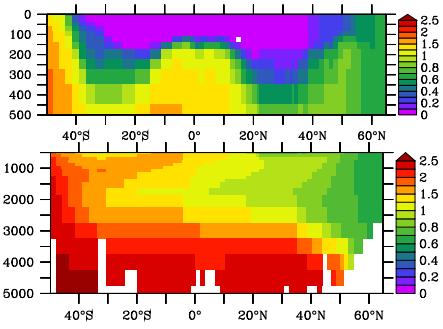
MITigsm, GA−02



PISCES 1, GA−02



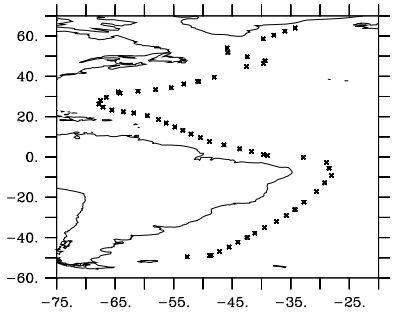
PISCES 2, GA−02

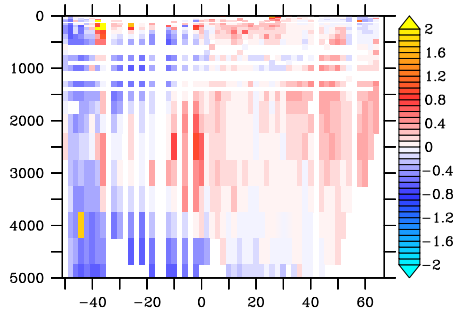


RECOM, GA−02

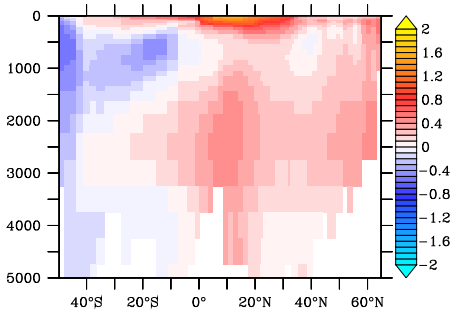
TOPAZ, GA−02

Festar based on Nitrate, GA−02, Cruise

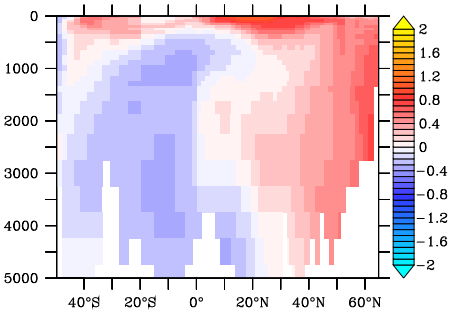




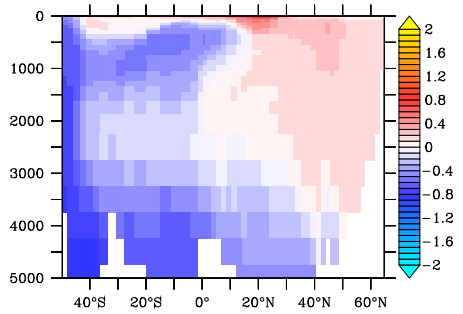
BEC



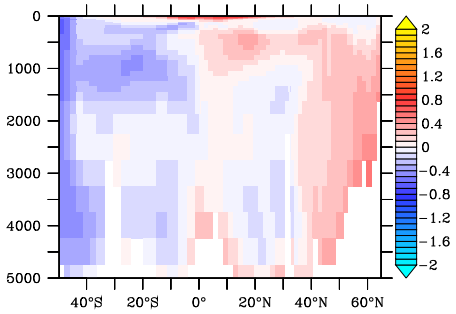
BFM



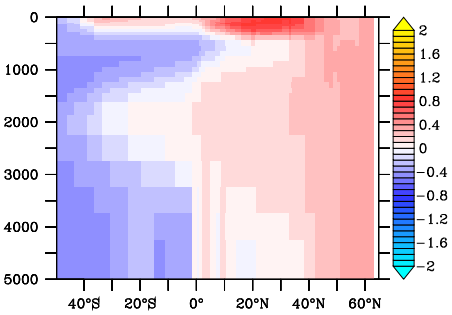
BLING



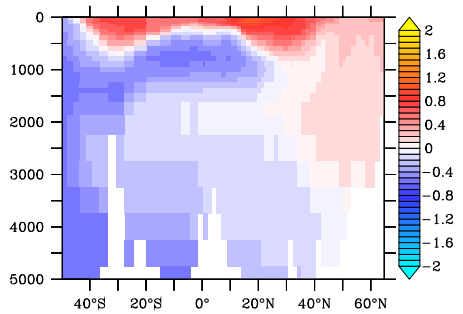
COBALT



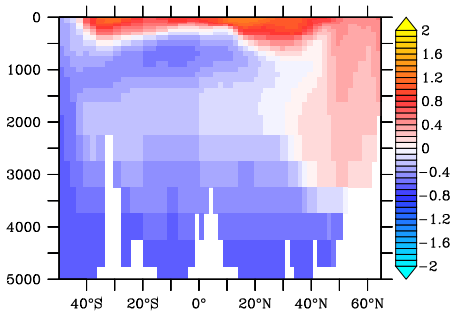
GENIE



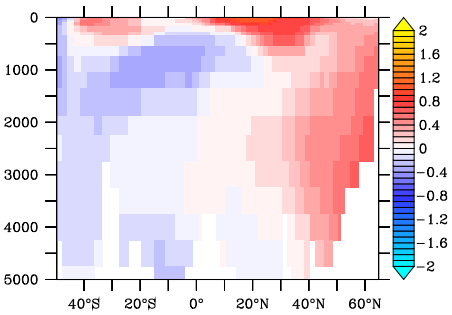
MEDUSA1



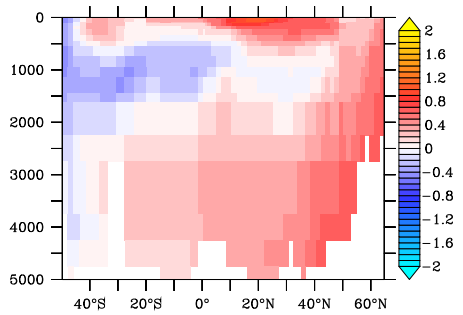
MEDUSA2



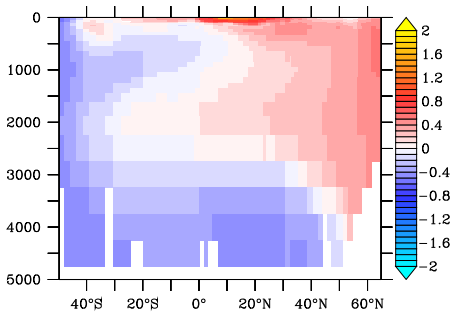
MITecco



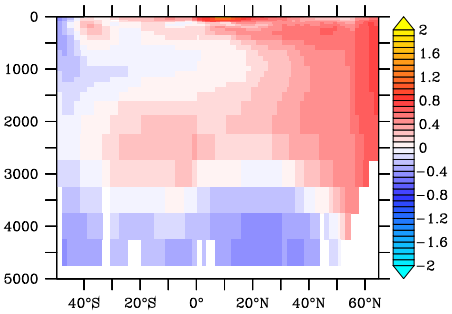
MITigsm



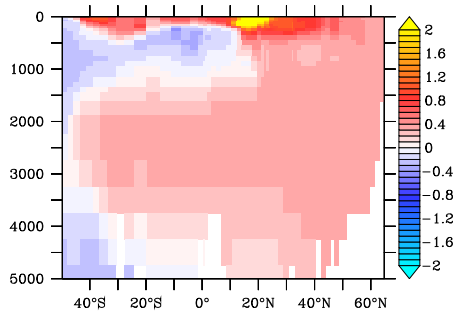
PISCES1



PISCES2



RECOM



TOPAZ

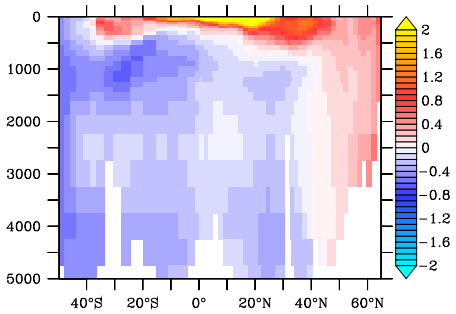


Table 1 A summary of the FeMIP models. Indicated is the number of years of model spin up, which iron sources are represented, whether ligands are present, fixed, or dynamic, whether Fe chemistry is consider implicitly (i.e. a threshold) or explicitly (i.e. computing free Fe as a function of ligands and conditional stability of complexes), the order of Fe scavenging (1 = uniform rate, 2 = also a function of particles), whether colloidal pumping loss of DFe is represented, if biological have a fixed or variable demand for Fe (Fe quota), if recycling is a fixed rate of variable (as a function of their Fe demand), how many particulate Fe pools are represented (if any) and whether the regeneration efficiency of particulate Fe is specific or is coupled to other tracers (carbon or nitrogen for example).

		Fe sources				**Fe chemistry**				**Fe biology**		**Particles**	
Model	**Spin Up**	**Dust**	**Sed.**	**Hydro.**	**River**	**Ligands**	**Speciation**	**Scav**	**Colloids**	**Demand**	**Recycling**	**Pools**	**Regen.**
BEC	290 yrs	yes	yes	yes	yes	Fixed	Implicit	2	no	Variable	Fixed	0	Coupled
BFM	30 yrs	yes	no	no	yes	Fixed	Explicit	1	no	Variable	Variable	1	Coupled
BLING	1,800 yrs	yes	yes	no	no	Fixed	Explicit	2	no	Variable	Fixed	1	Coupled
COBALT	100 yrs	yes	yes	no	no	Fixed	Explicit	1	no	Variable	Variable	1	Specific
GENIE	10,000 yrs	yes	no	no	no	Fixed	Explicit	2	no	Variable	Variable	1	Coupled
MEDUSA1	40 yrs	yes	no	no	no	Fixed	Explicit	1	no	Fixed	Fixed	1	Coupled
MEDUSA2	140 yrs	yes	yes	no	no	Fixed	Explicit	1	no	Fixed	Fixed	1	Coupled
MITecco	40 yrs	yes	yes	no	no	Fixed	Explicit	2	no	Fixed	Variable	1	Coupled
MITigsm	190 yrs	yes	yes	no	no	Fixed	Explicit	2	no	Fixed	Variable	1	Coupled
PISCES1	3,000 yrs	yes	yes	yes	yes	Fixed	Explicit	2	yes	Variable	Variable	2	Coupled
PISCES2	3,000 yrs	yes	yes	yes	yes	Dynamic	Explicit	2	yes	Variable	Variable	2	Coupled
REcoM	1,000 yrs	yes	yes	no	no	Fixed	Explicit	2	no	Both	Fixed	1	Coupled
TOPAZ	1,000 yrs	yes	yes	no	no	Dynamic	Explicit	2	no	Variable	Variable	1	Coupled

Table 2. A summary of the magnitude of the Fe sources, the total and average Fe inventories, and the residence time of Fe across the FeMIP models.

	Fe sources Gmol yr^{-1}							
Model	**Dust**	**Sed**	**Hydro**	**Rivers**	**Total**	**Fe inventory x10^{11} mol**	**Average Fe (nmoles L^{-1})**	**Residence Time (yrs)**
BEC	21.9	84.6	17.7	0.34	124.5	10.1	0.74	8.1
BFM	1.4	0	0	0.06	1.4	8.8	0.65	626.3
BLING	3.3	9.1	0	0	12.4	5.3	0.37	42.4
COBALT	32.5	155	0	0	182.5	6.8	0.50	3.7
GENIE	1.8	0	0	0	1.8	10.1	0.48	560.0
MEDUSA1	2.7	0	0	0	2.7	6.3	0.46	232.0
MEDUSA2	3.4	2.9	0	0	6.8	4.8	0.35	69.9
MITecco	3.5	104	0	0	107.5	8.8	0.65	8.2
MITigsm	1.4	194	0	0	195.4	9.0	0.66	4.6
PISCES1	32.7	26.6	11.3	2.5	71.0	8.1	0.59	11.5
PISCES2	32.7	26.6	11.3	2.5	71.0	11.2	0.81	15.7
REcoM	3.7	0.6	0	0	4.3	12.5	0.73	291.6
TOPAZ	13.8	74.8	0	0	88.6	6.8	0.50	7.6
				Mean	66.9	8.3	0.58	144.7
				St dev	67.1	2.2	0.14	175.8

Table 3. Correlation coefficient (R) and in parentheses the mean bias (nM) between the different FeMIP models and the expanded database of Tagliabue et al. [2012] across different depth bins. Average dissolved iron data (nM) for the different depth strata is presented in the final row. The iron data are gridded on the FeMIP grid as described in the text.

	ALL	**0-100**	**100-500**	**500-1000**	**2000-5000**
Model					
BEC	0.51 (-0.02)	0.48 (0.23)	0.52 (-0.05)	0.47 (-0.15)	0.31 (0.01)
BFM	0.39 (-0.48)	0.34 (-0.29)	0.36 (-0.47)	0.33 (-0.52)	-0.03 (-0.48)
BLING	0.37 (-0.33)	0.37 (-0.13)	0.49 (-0.17)	0.46 (-0.26)	0.01 (-0.44)
COBALT	0.45 (-0.25)	0.38 (-0.19)	0.48 (-0.25)	0.51 (-0.19)	-0.11 (-0.25)
GENIE	0.25 (-0.28)	0.43 (-0.02)	0.46 (-0.11)	0.43 (-0.20)	-0.14 (-0.40)
MEDUSA1	-0.01 (-0.24)	0.37 (0.23)	0.38 (0.04)	0.07 (-0.24)	0.07 (-0.37)
MEDUSA2	-0.14 (-0.32)	0.35 (0.29)	0.37 (0.07)	-0.06 (-0.30)	-0.10 (-0.51)
MITecco	0.39 (-0.12)	0.34 (-0.10)	0.36 (-0.09)	0.33 (-0.14)	-0.03 (-0.11)
MITigsm	0.37 (-0.14)	0.04 (-0.22)	0.42 (-0.24)	0.29 (0-.24)	-0.13 (-0.04)
PISCES1	0.47 (-0.23)	0.36 (-0.06)	0.47 (-0.17)	0.47 (-0.03)	0.21 (-0.27)
PISCES2	0.51 (-0.04)	0.37 (0.03)	0.52 (0.01)	0.43 (-0.03)	0.35 (-0.05)
REcoM	0.39 (-0.05)	0.33 (0.25)	0.40 (0.01)	0.44 (-0.01)	-0.04 (-0.12)
TOPAZ	0.10 (-0.13)	0.42 (0.67)	0.27 (0.26)	0.33 (-0.12)	0.01 (-0.34)
Data	0.64	0.52	0.63	0.76	0.90